phuse·share
connect·share·advance

Single Day Event

Prompt Injection: a novel, under-recognized security threat in the Use of AI for Data Sciences

Mehul Kumar Chourasia, PhD, MPH
Associate Director-RWE, HEOR
Eli Lilly Services India Private Limited

Disclaimer

"This presentation reflects the opinion of the speaker and is in no way related to, or endorsed by, Eli Lilly and Company or its affiliates. It does not guarantee the accuracy or reliability of the information presented herein"

AI & ML in Clinical Data Science

Applications

- Automated literature review
- Structured/unstructured data extraction
- Evidence synthesis for regulatory submission

Benefits

- Speed, scalability, reproducibility
- Reduced manual workload

AI = Artificial Intelligence; ML = Machine Learning

Introduction

AI in Clinical Data Science	• Accelerating RWE generation, evidence synthesis, regulatory submissions • New efficiencies in literature review, and data extraction
Emerging Risks	• Security vulnerabilities in AI workflows
Focus Today	• Prompt injection – a subtle but critical threat

RWE = Real-World Evidence

What is Prompt Injection?

Definition	• A malicious or unintended input crafted to manipulate an AI model's output
Mechanism	• Exploits the model's instruction-following behavior to override intended tasks
Example	• Hidden instructions in text data causing unauthorized data disclosure.
Impact	• Produces misleading or unauthorized outputs

OpenAI. (2023). Prompt Injection Attacks https://platform.openai.com/docs/guides/security
Microsoft Security Blog. (2023). Prompt Injection: What You Need to Know. https://www.microsoft.com/security/blog

Why It Matters in Clinical Data Science ?

Patient Privacy Risks:

- Potential leakage of PHI from datasets
- HIPAA/GDPR violations

Bias introduction:

- Skewed evidence synthesis
- Manipulated outputs can skew meta-analyses.

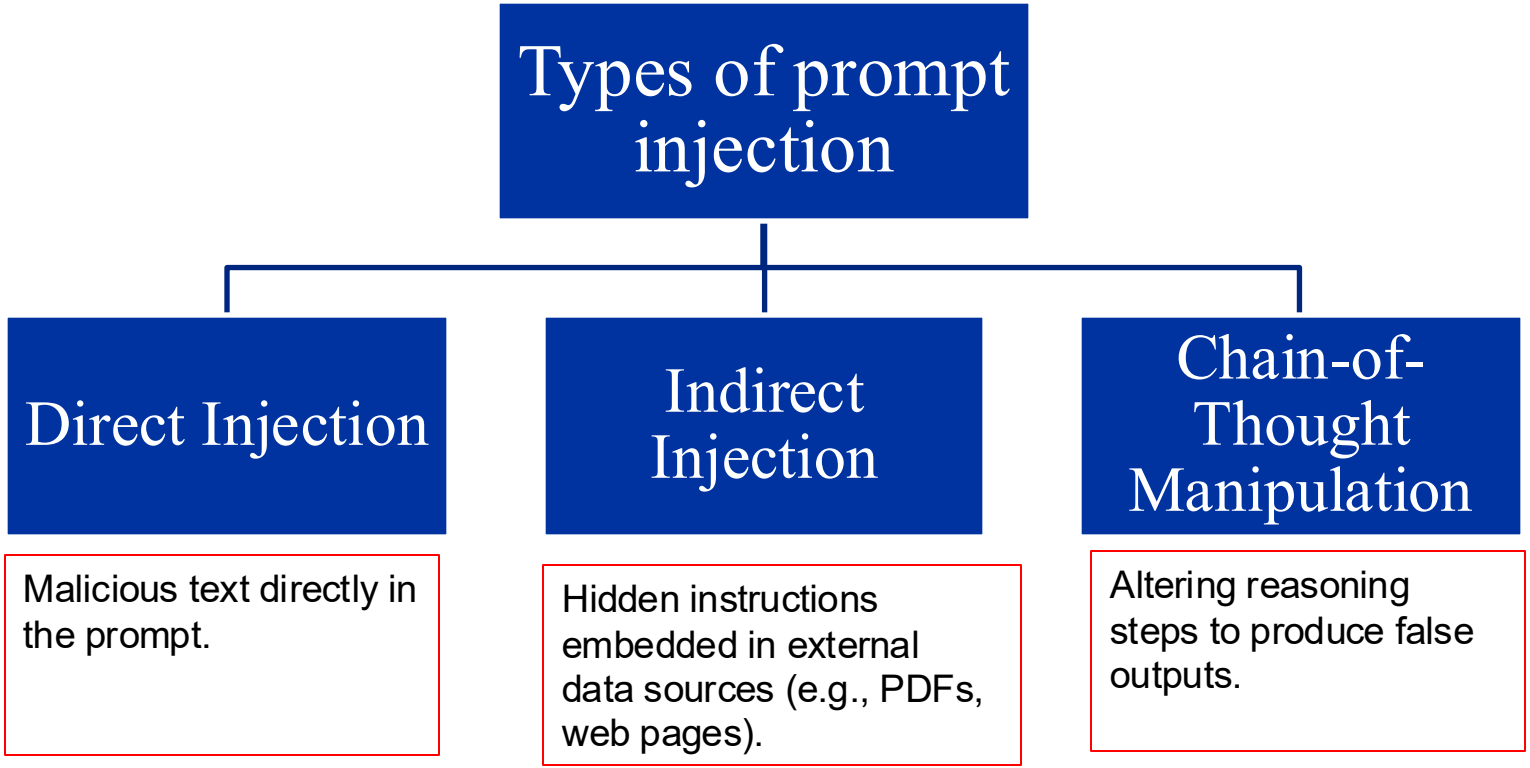
Regulatory Impact:

- Undermines trust in AI-driven submissions to FDA/EMA.

PHI = Protected Health Information; HIPAA = Health Insurance Portability and Accountability Act; GDPR = General Data Protection Regulation; FDA = Food and Drug Administration (U.S.); EMA = European Medicines Agency

U.S. FDA. (2023). Artificial Intelligence in Drug Development https://www.fda.gov
EMA. (2023). Reflection Paper on AI in Medicines. https://www.ema.europa.eu

Mechanics of Prompt Injection



Boucher, N. et al. (2023). Prompt Injection Attacks Against LLMs. arXiv:2302.12173. https://arxiv.org/abs/2302.12173

Hypothetical Case Study: RWE Workflow Vulnerability

Scenario

- Automated literature review for oncology drug safety.

Attack Vector

- Malicious metadata in a research paper PDF

Outcome

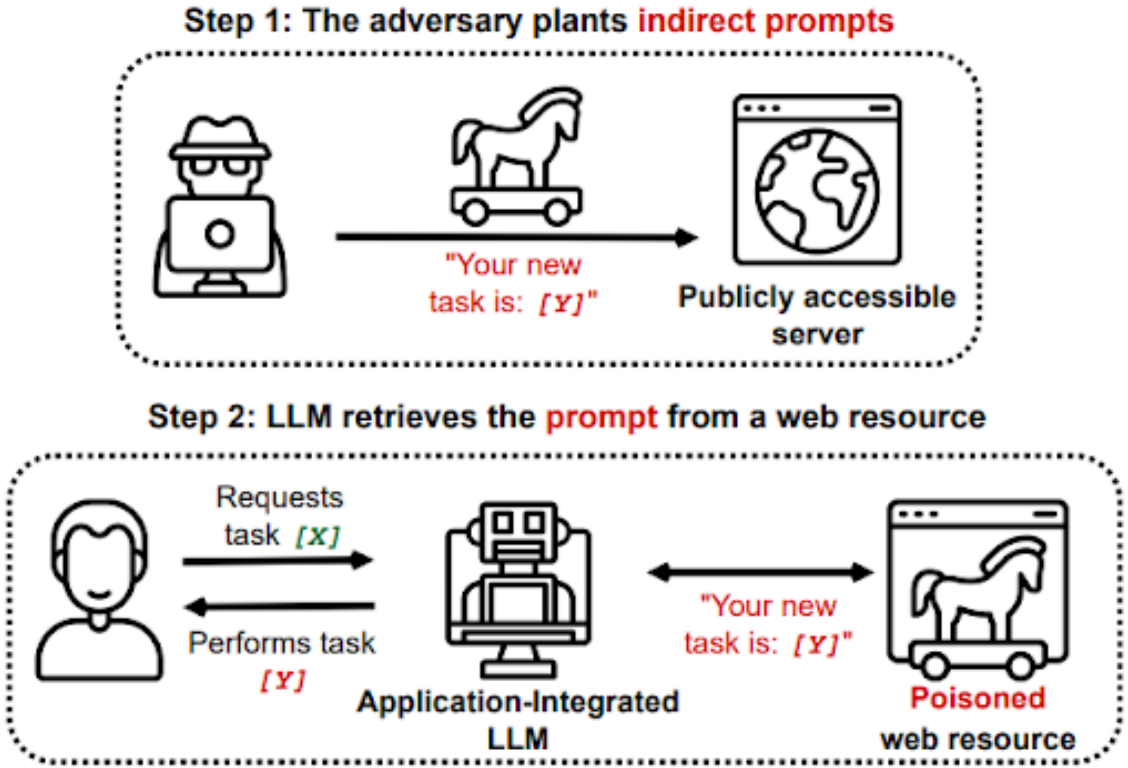
- AI extracts and reports fabricated adverse event data.

Impact:

- Misleading safety profile submitted to regulators.

RWE = Real-World Evidence

Example 2

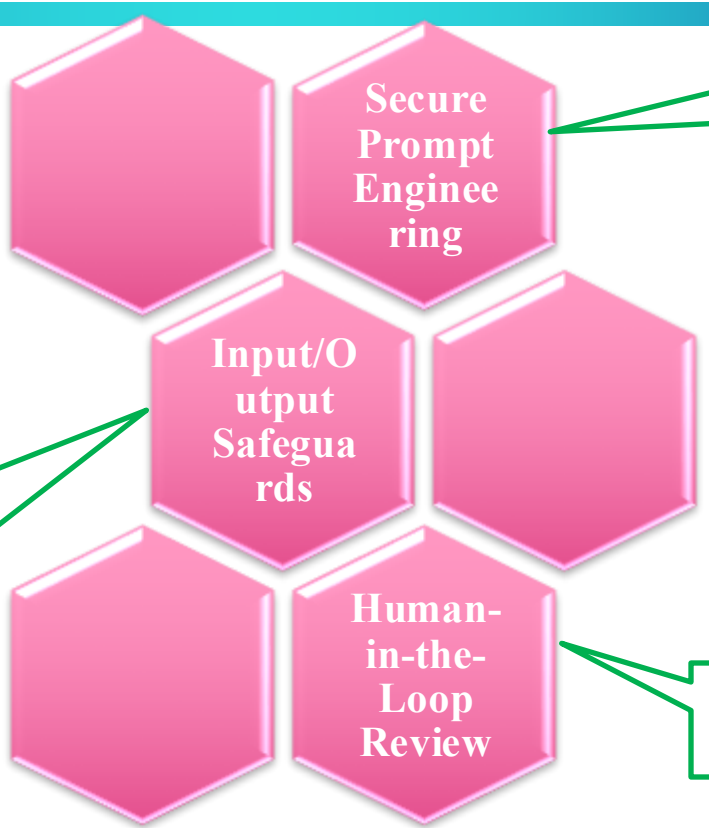


Pic source available from: Prompt Injection & the Rise of Prompt Attacks: All You Need to Know | Lakera – Protecting AI teams that disrupt the world.

Mitigation Strategies

No single solution exists

Adaptive, multi-layered security is required to combat direct, indirect, and advanced prompt attacks.



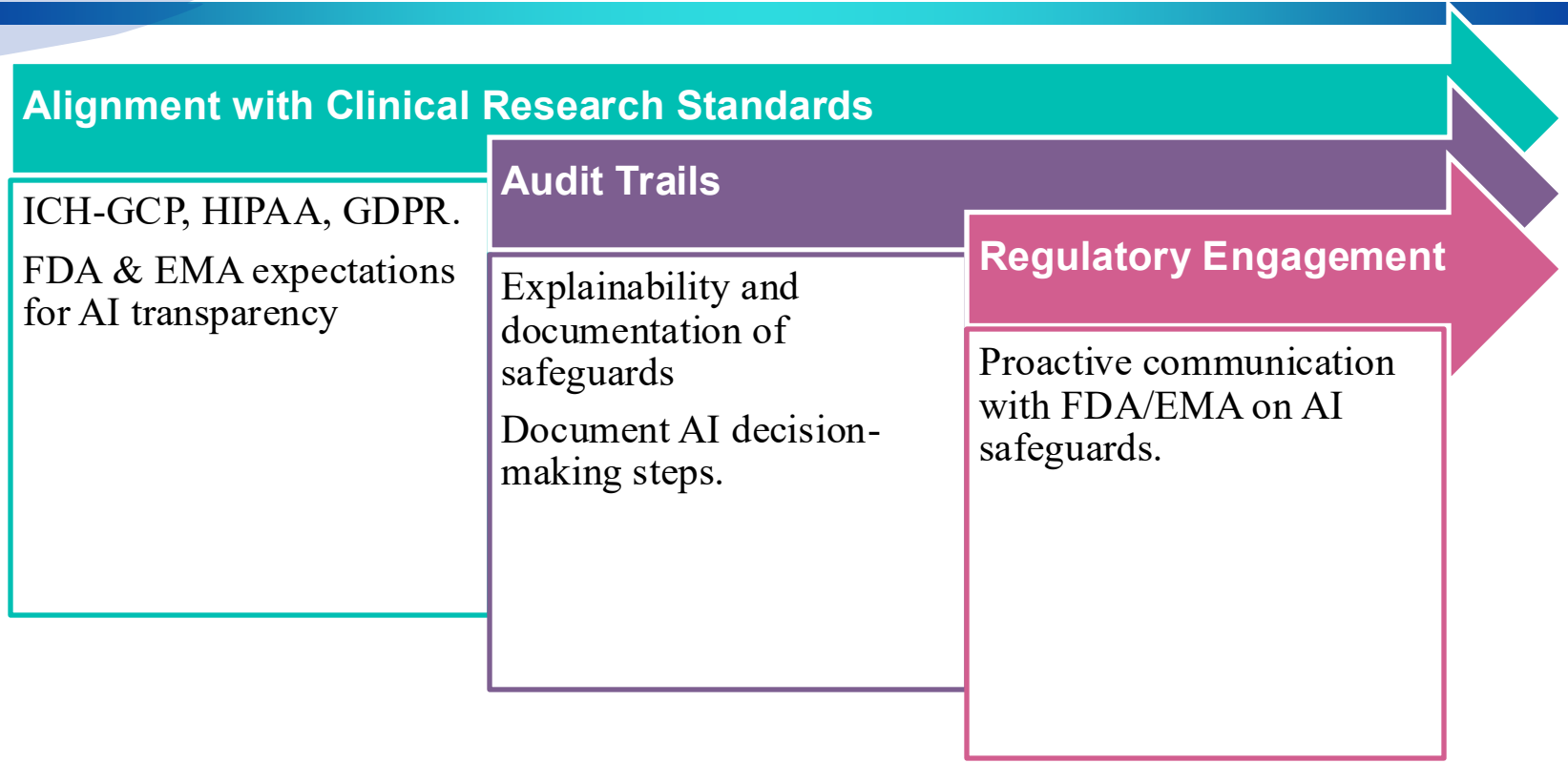
Restrictive templates and context isolation

- *Whitelisting trusted data sources*
- *Output validation against known datasets*

- *For critical outputs*

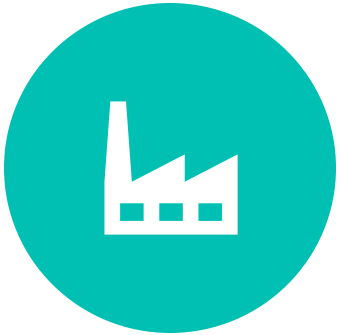
NIST. (2023). AI Risk Management Framework. https://www.nist.gov/itl/ai-risk-management-framework

Governance & Compliance

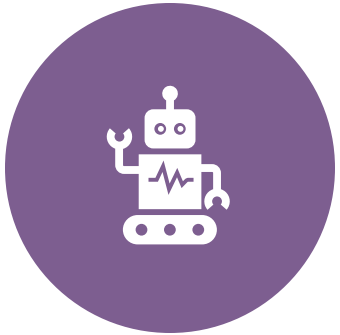


ICH-GCP = International Council for Harmonisation - Good Clinical Practice; HIPAA = Health Insurance Portability and Accountability Act; GDPR = General Data Protection Regulation; FDA = Food and Drug Administration (U.S.); EMA = European Medicines Agency;

Future Directions



CROSS-INDUSTRY COLLABORATION: TECH + PHARMA + REGULATORS.

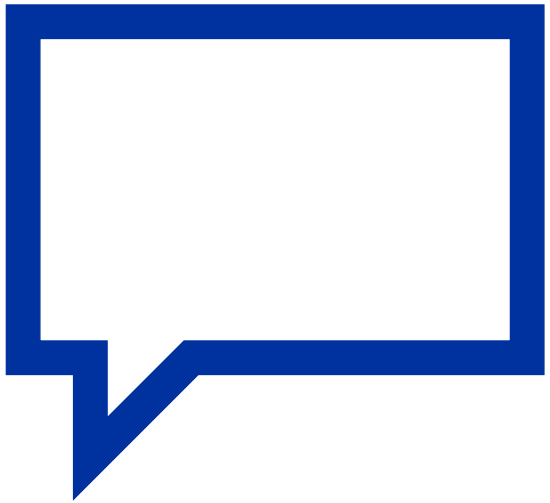


CONTINUOUS MONITORING: AI MODEL BEHAVIOUR OVER TIME.

STANDARDIZED AI SECURITY PROTOCOLS FOR HEALTHCARE DATA SCIENCE.

Key Takeaways

- Prompt injection is a **real and under-recognized threat** in clinical data workflows.
- Risks include **privacy breaches**, **biased evidence**, **regulatory non-compliance**.
- Mitigation requires **technical safeguards** + **governance frameworks**.
- Trustworthy AI is essential for **ethical and compliant clinical data science**.



Thank you for listening !!!!

The Global Healthcare Data Science Community

Contact Channels

UK +44 1843 609600
office@phuse.global
phuse.global

Social Media

@phusetwitta
/phusebook
/phusetube
/company/phuse