DevSecOps for Enhancing Security for Machine Learning

Do ML, the DevOps way!

Rajendra Prasad (RP), Aditi Kulkarni, Vijeth Hegde

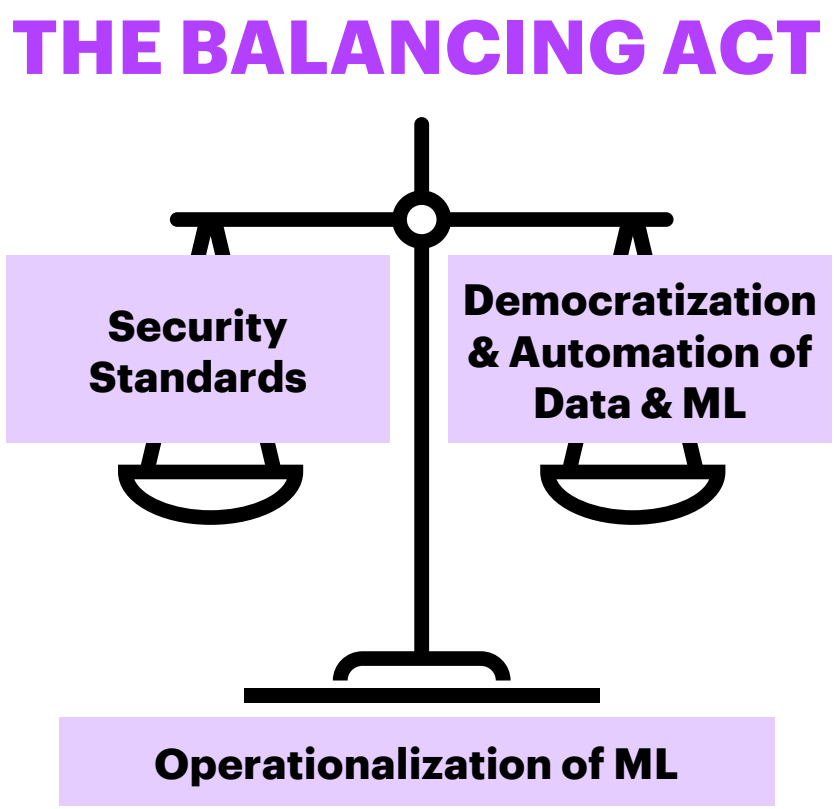
accenture

December 2021

Why do we need security for ML?

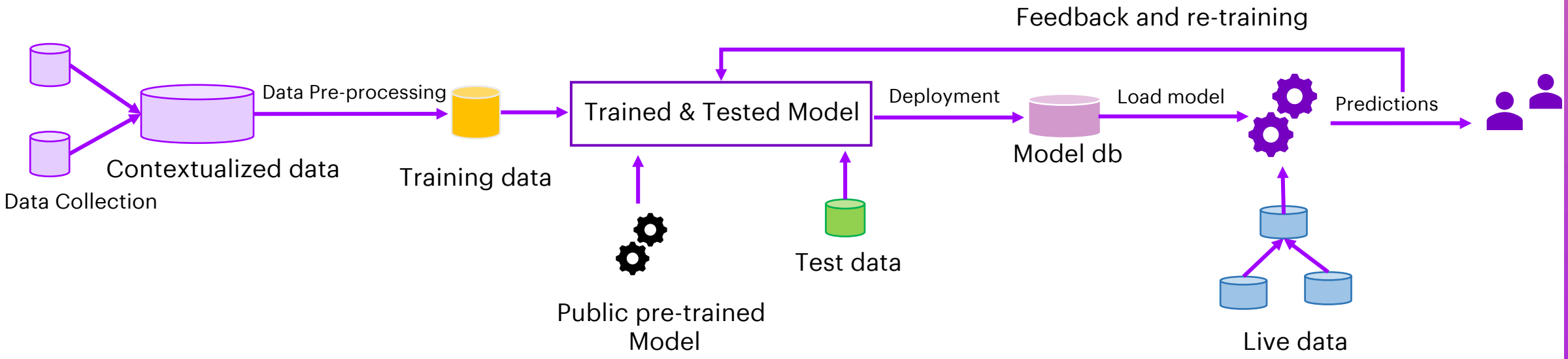
- **Data** is the most important aspect of an ML system.
- The most important category of computer security risk is **malicious input**.
- ML will be **implied** in every software in the next 5 years.
- **Publicly** available models are extensively used, **transfer learning**.
- ML results are **poorly explained** and **impossible to reproduce** at times. When we can not reproduce and if nobody monitors the results then attacks can happen.

What it means for an organization

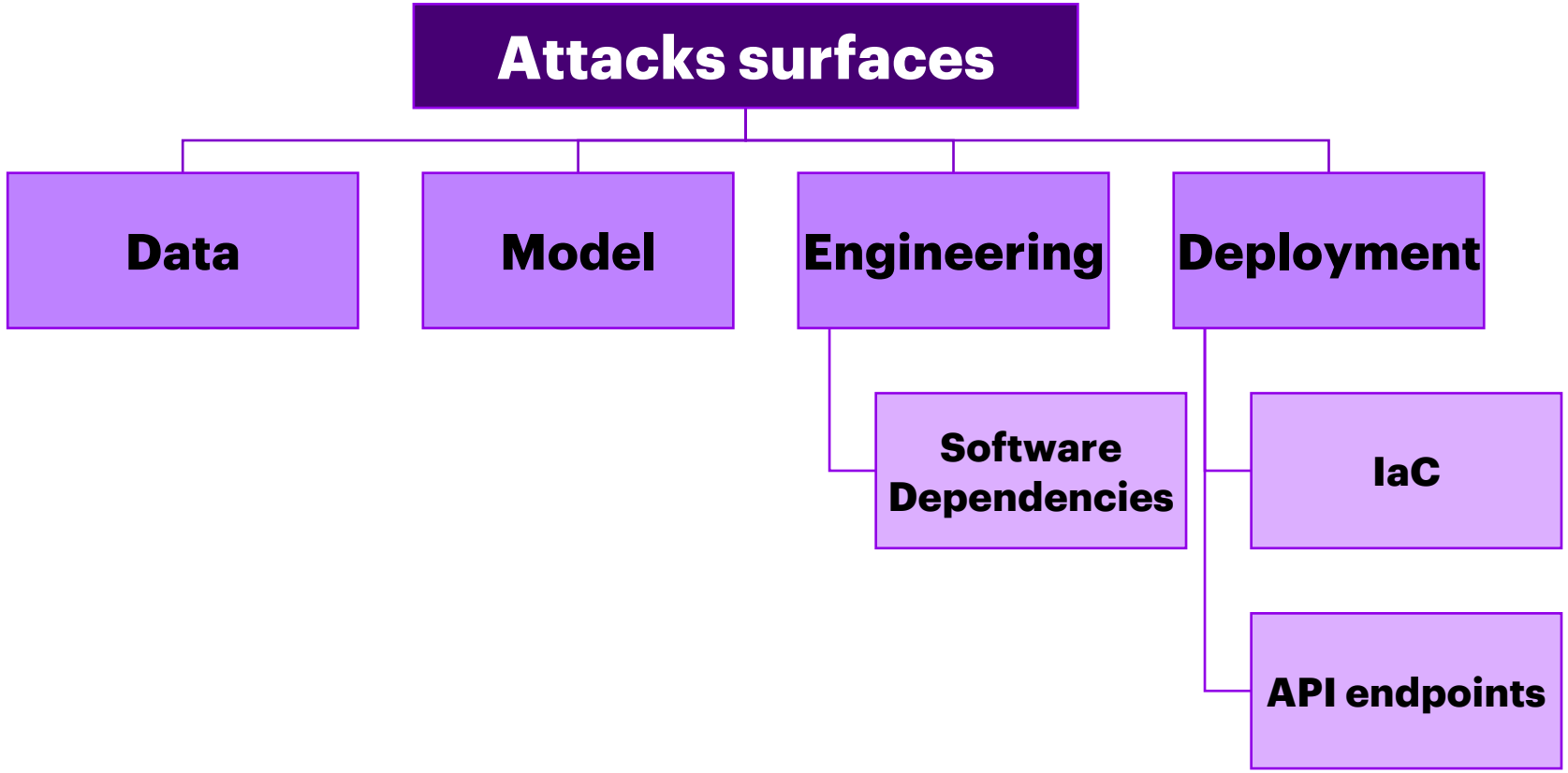


- Executives and CXOs need to do **“The Balancing Act”**.
- ML practitioners and engineers need to understand how “**shift-left security**” mentality can be applied to ML.
- Security practitioners need to understand **how ML will impact the security** of entire system.

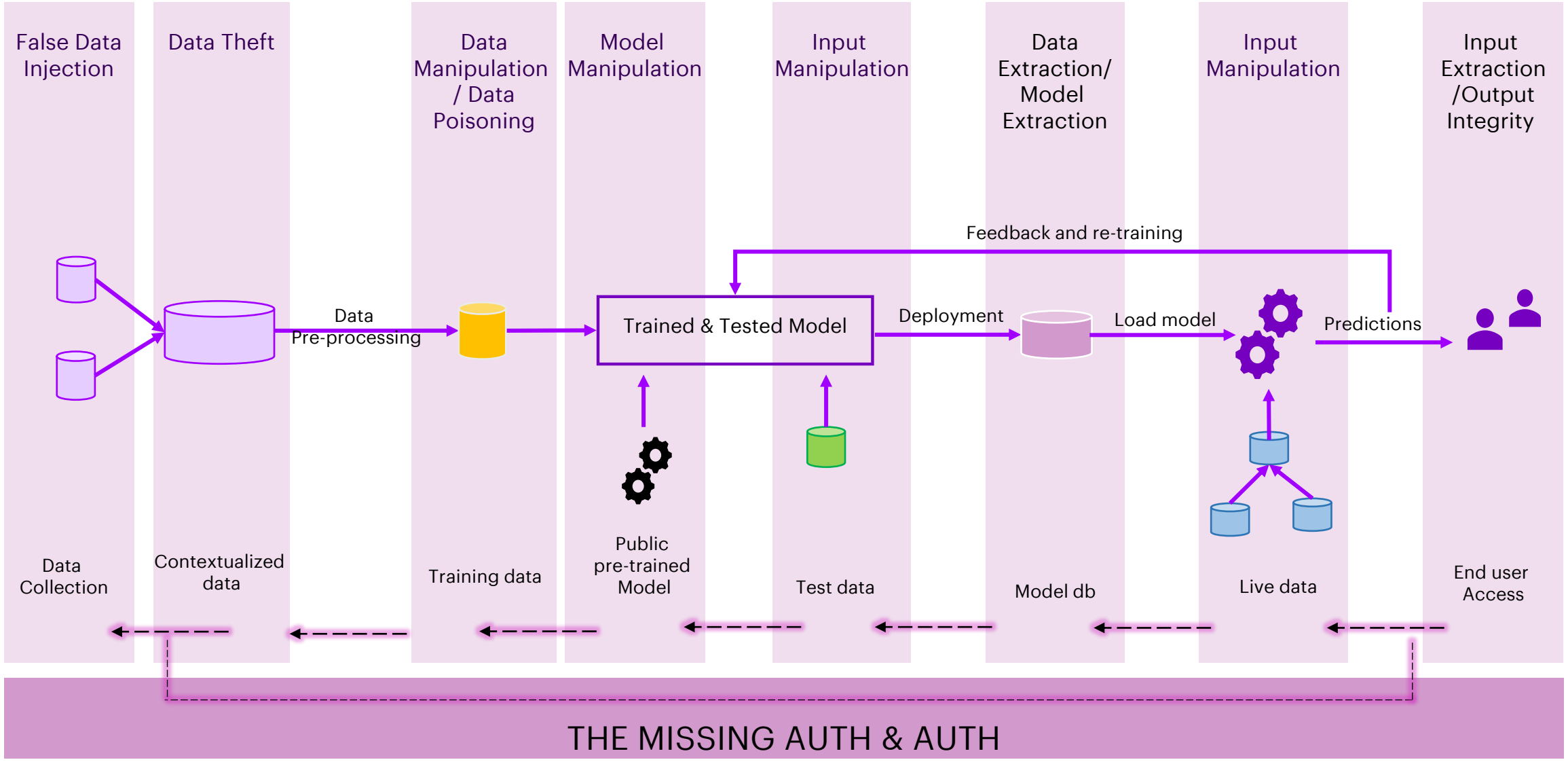
A typical AI/ML implementation



Common attack surfaces



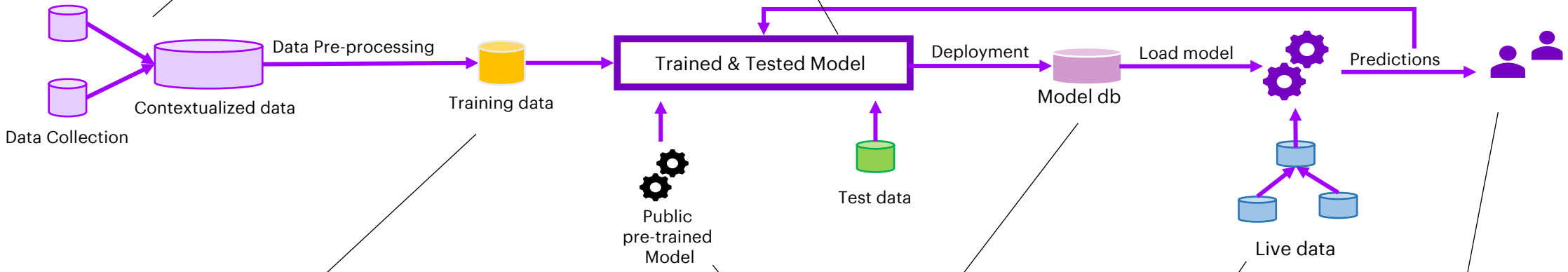
Common attacks on ML systems



A quick solution? The silo-ed defenses

Collect only the relevant data needed for training and predictions in production

RBAC and governance
Better algorithm choices



Anonymization, tokenization and encryption of data
Identify and remove PII data
Account for bias in the data

Model authentication at runtime
Model Hashing

Cryptography to harden the inputs to the models

Spam detection
Rate limiting
Authentication
Refresh the ML system in production to a known state, reset, or otherwise clean it periodically

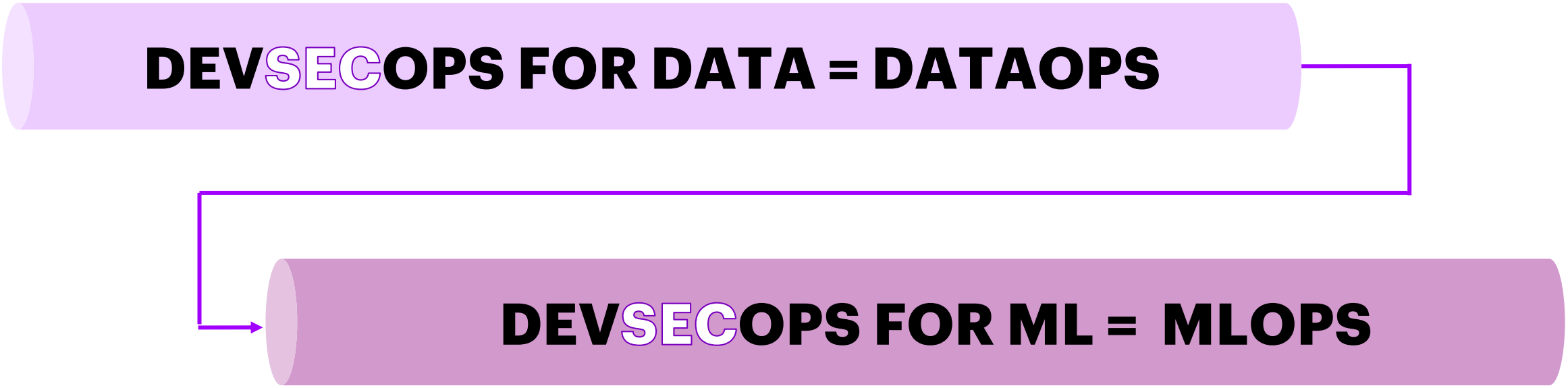
The two weakest links

DATA

AND

S . I . L . O . S

The unified defense



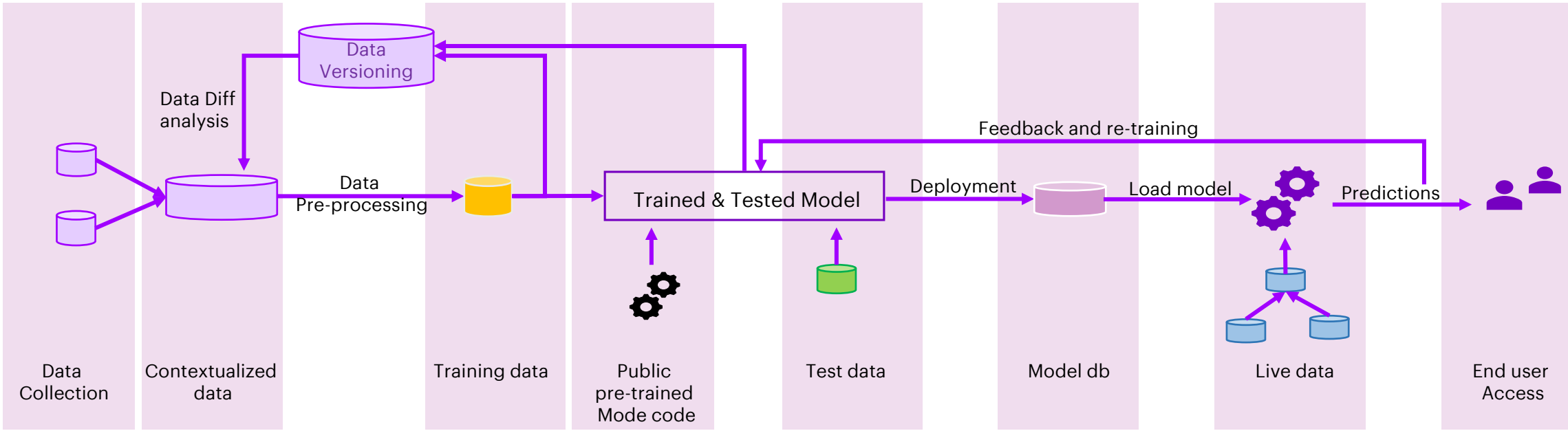
DATAOPS

- Data Quality
- Data Security
- Data Encryption

- Data Drift
- Data Lineages
- Data Versioning

- Model Versioning
- Model Drift

- Model Hashing
- Model Security



MLOPS

- Data Confidentiality
- Data Obfuscation
- Featurization of data
- DBOps

- Data Catalogues
- Loops in the data stream
- DataGovOps
- Outlier detection

- API Security
- Monitor outputs

- Refresh ML systems
- ML-SRE
- RBAC

DEVSECOPS FOUNDATION

AI-Driven | Shift Left security | Everything as Code (EaC) | Continuous Data & ML Engineering
CI | CT | CS | CD | CF | Code to Cloud Visibility
Regular Software patching | Scans for External libraries

Principle of Least Privilege | Principle of Defense-in-depth | Secure the Weakest Link

SECURITY AS A CULTURE

Do ML, the DevOps way !

1. Everything as Code – Automate Everything
2. Shift-Left Security
3. Orchestrate across teams
4. Fail securely | Fail cheaply | Fail fast
5. Innovate Fearlessly

Thank You!