AI operationalization

Vincent Nelis, PhD
Manager, Senior Data Scientist

Main pain-points in AI operationalization

Based on my own experience

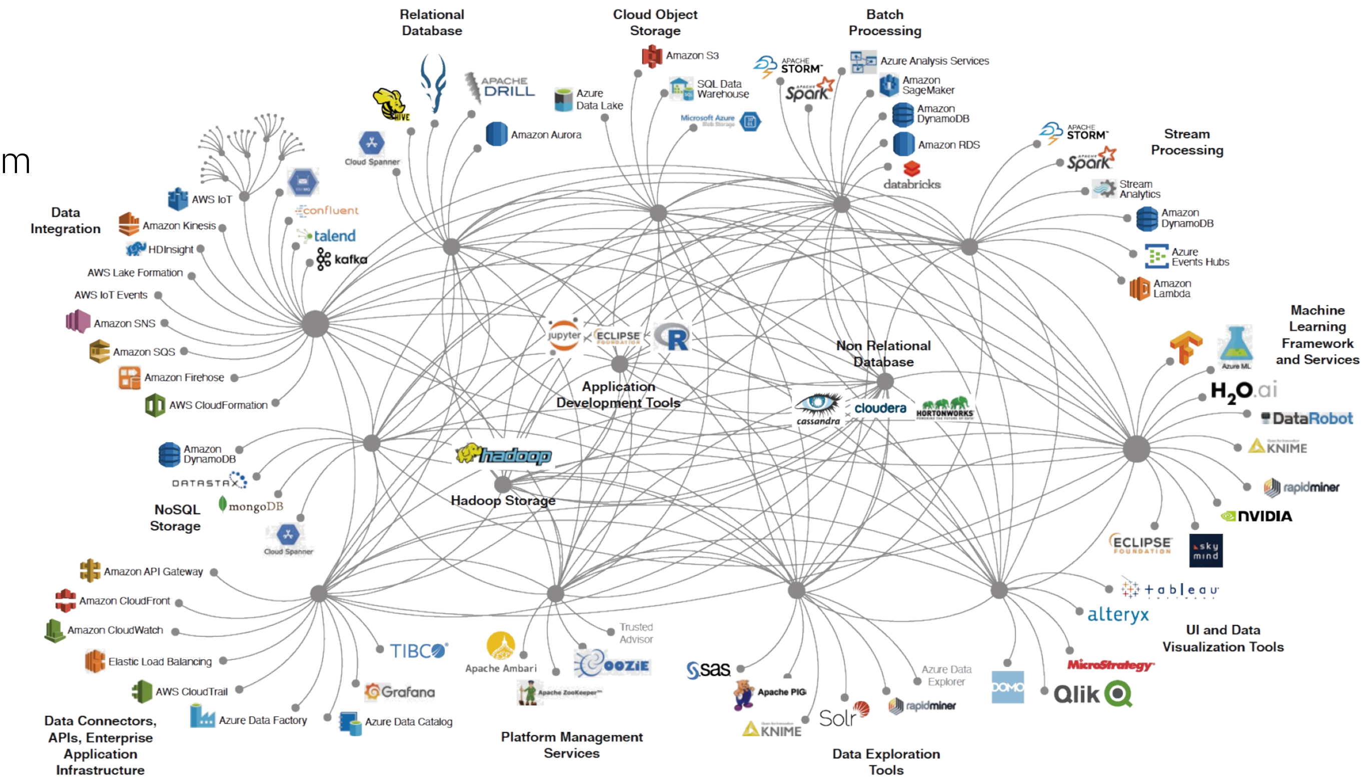
1. *Infrastructure* — Why should we pay for something that we can build ourself for free?

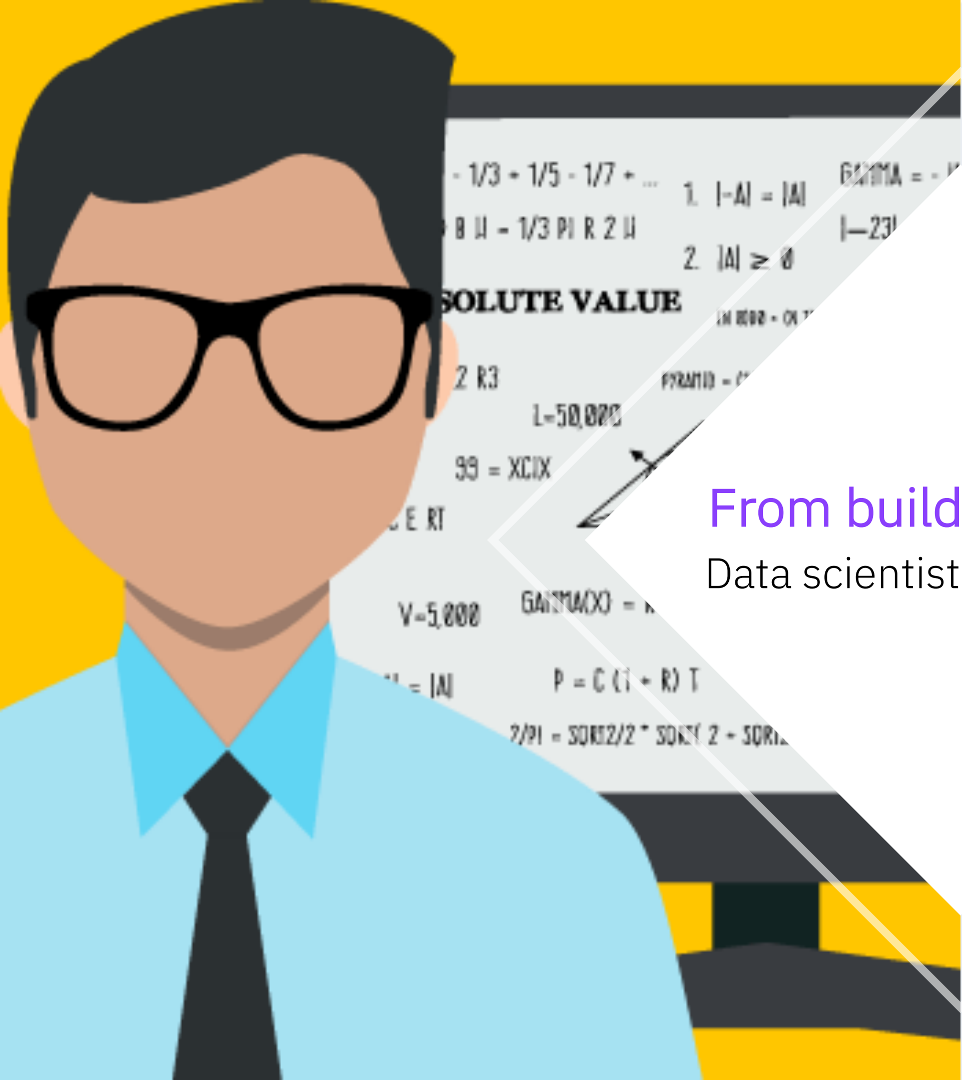
2. *Deployment* — A model has been trained and business value has been demonstrated. How can we now deploy it to production?

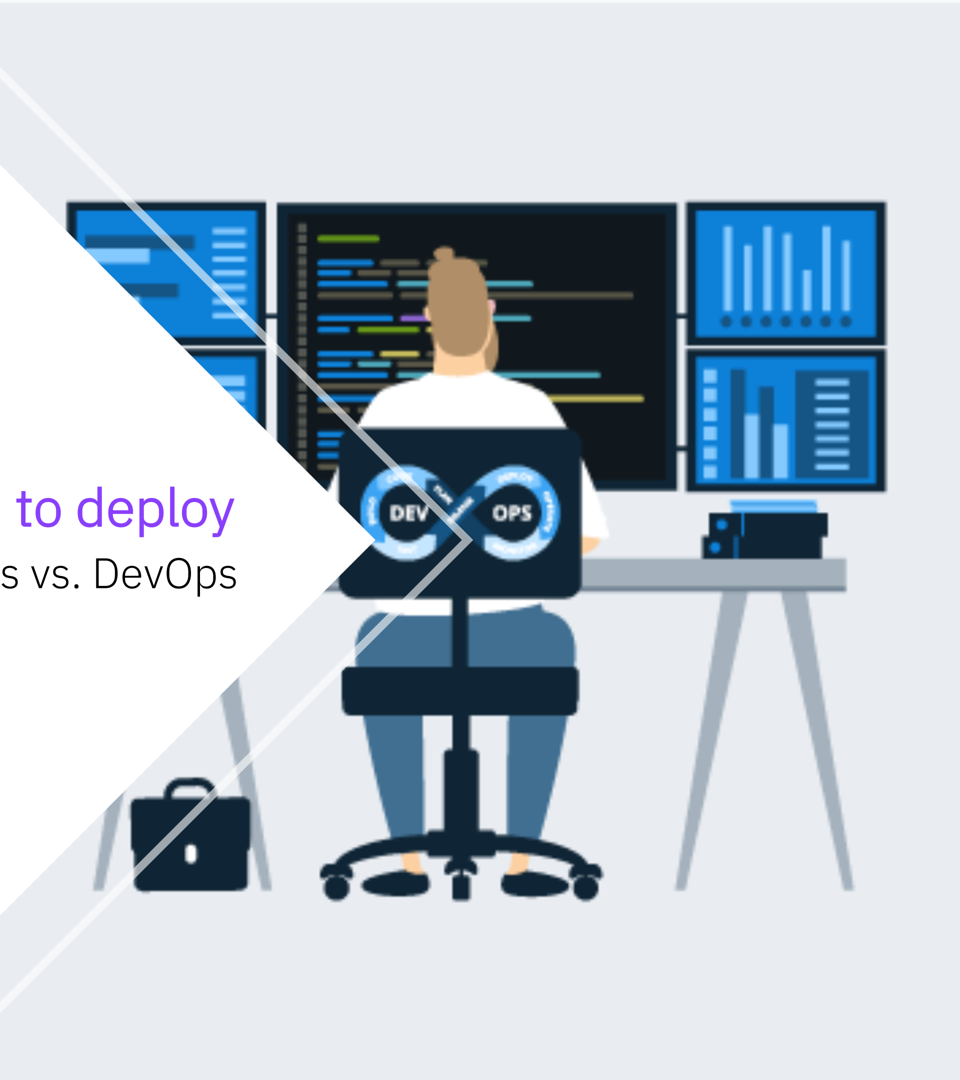
3. *Operation* — The model has been deployed. How can we trust its output and make sure it is “safe” to use?

Infrastructure: The “build it yourself” approach

The “build it yourself” approach requires stitching together dozens of disparate open source components from different developers with different APIs, different code bases, and different levels of maturity and support. These components were not designed to work together, resulting in a degree of complexity that overwhelms even the best development teams.







From build to deploy

Data scientists vs. DevOps

Operation: What is a ML model?

How can we trust that thing?

Facts:

1. The model is a mathematical function that has been tuned to fit a certain dataset called the training data
2. The same model is now used to predict values from newly generated data, potentially data seen for the very first time.
3. The output of the model has a built-in error from day 1.
4. The error is growing over time (data drift)

Operation: The five pillars of trustworthy AI

Brand new challenges

1. *Transparent* — Transparency reinforces trust and sharing information with stakeholders of varying roles engenders trust.

2. *Explainable* — How AI-led decisions are made and what determining factors were included are crucial to understand.

3. *Fair* — Ensuring proper monitoring and safeguards are in place to mitigate bias and drift leads to fairer treatment for all.

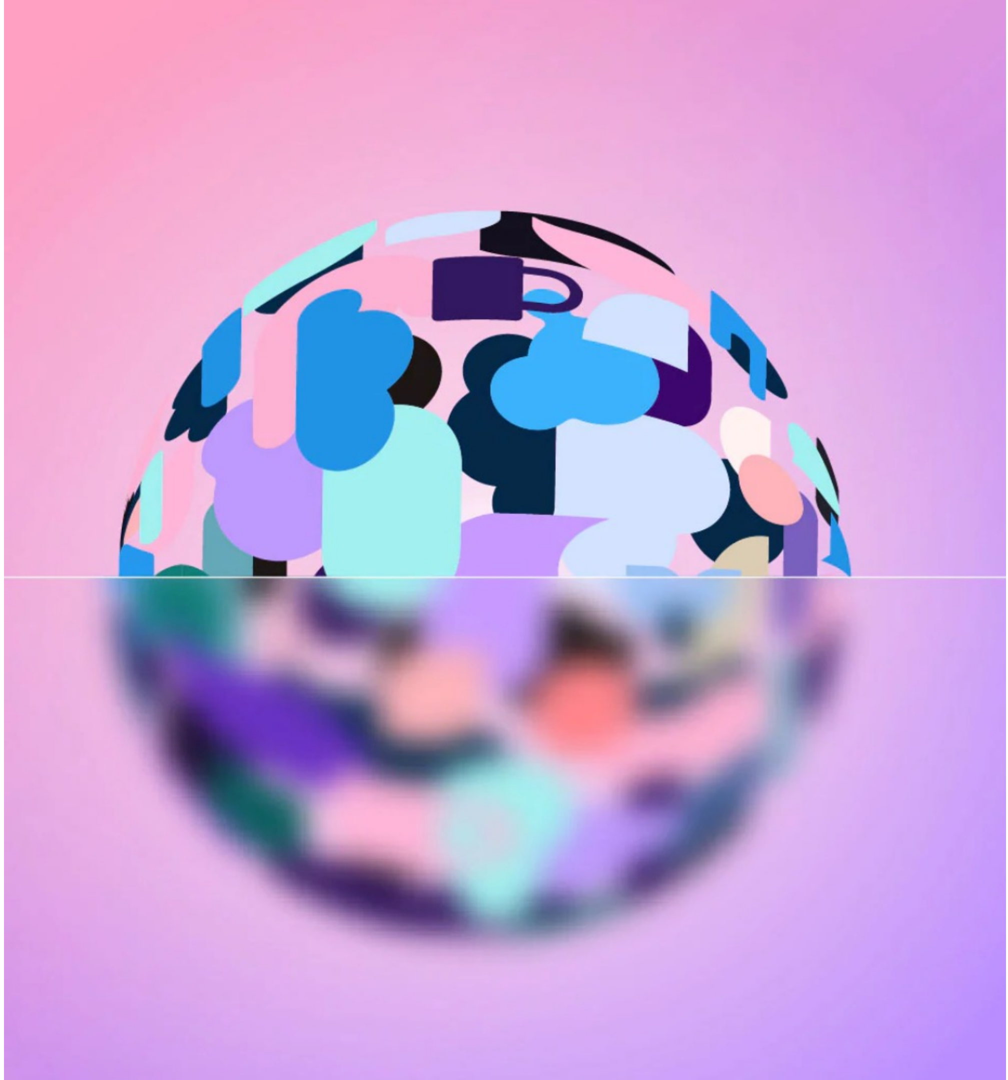
4. *Robust* — Guarding against adversarial threats and potential incursions to keep systems healthy.

5. *Privacy* — AI systems safeguard data through the entire lifecycle from training to production and governance.

Trustworthy AI:

The 5 pillars

1. Transparent
2. Explainable
3. Fair
4. Robust
5. Privacy



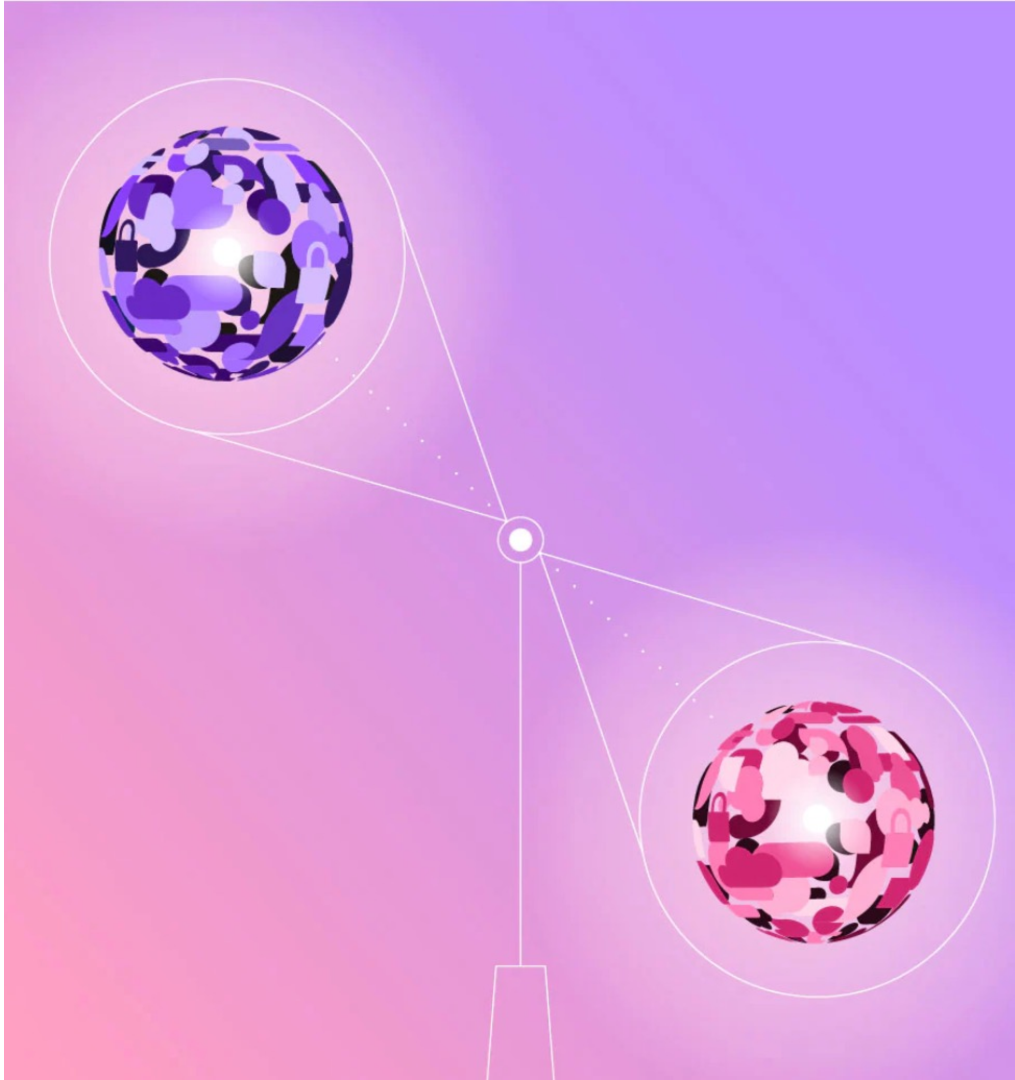
Trustworthy AI:
The 5 pillars

1. Transparent
2. Explainable
3. Fair
4. Robust
5. Privacy

Trustworthy AI:

The 5 pillars

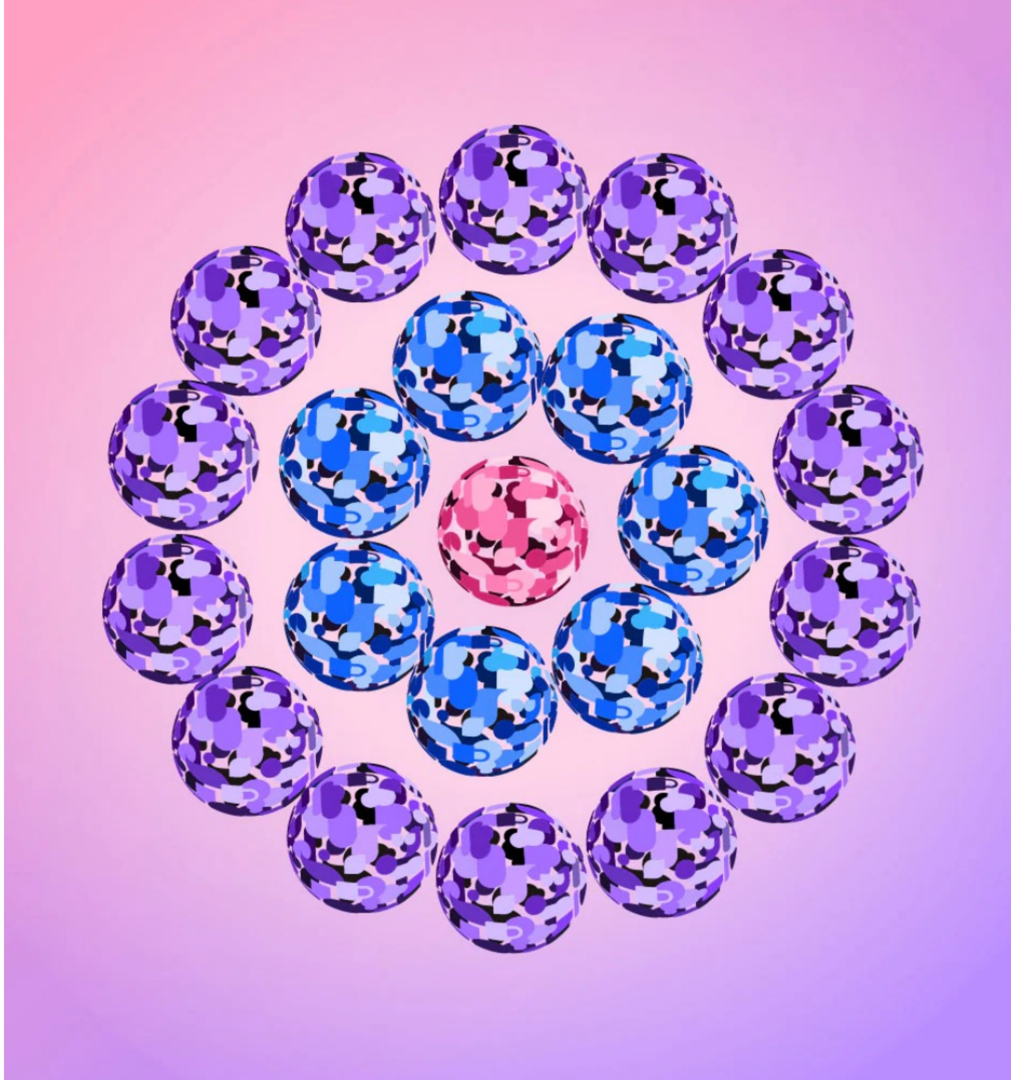
1. Transparent
2. Explainable
3. Fair
4. Robust
5. Privacy



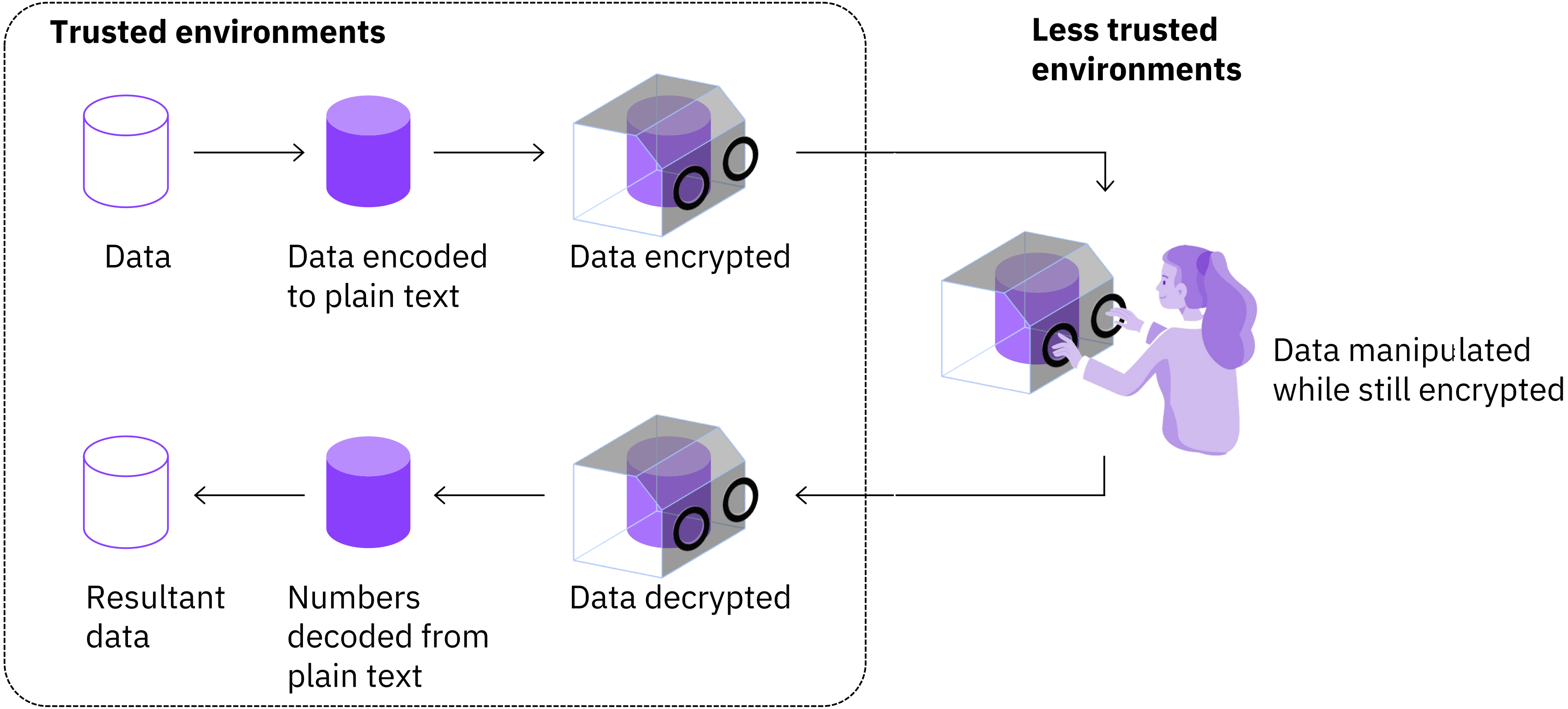
Trustworthy AI:

The 5 pillars

1. Transparent
2. Explainable
3. Fair
4. Robust
5. Private



Fully homomorphic encryption


Trusted environments
Data
Data encoded to plain text
Data encrypted
Less trusted environments
Data manipulated while still encrypted
Resultant data
Numbers decoded from plain text
Data decrypted

Trustworthy AI:
The 5 pillars

1. Transparent
2. Explainable
3. Fair
4. Robust
5. **Privacy**

Vincent Nelis, PhD
Manager
Senior Data Scientist
IBM Data Science & AI Elite Team
vincent.nelis@ibm.com

Advance trustworthy AI
from *principle* to **practice**

ibm.com/watson/trustworthy-ai