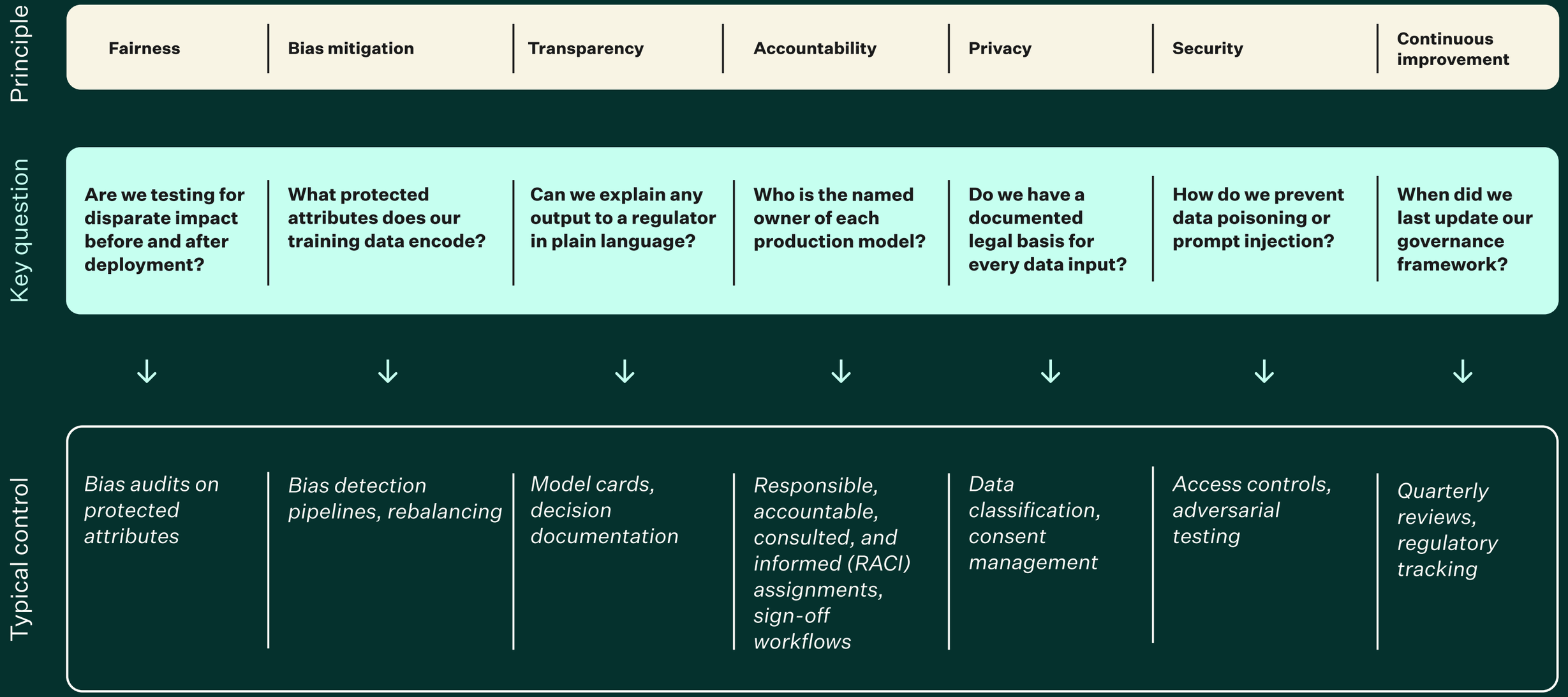
Principle
Fairness
Bias mitigation
Transparency
Accountability
Privacy
Security
Continuous improvement
Key question
Are we testing for disparate impact before and after deployment?
What protected attributes does our training data encode?
Can we explain any output to a regulator in plain language?
Who is the named owner of each production model?
Do we have a documented legal basis for every data input?
How do we prevent data poisoning or prompt injection?
When did we last update our governance framework?
Typical control
Bias audits on protected attributes
Bias detection pipelines, rebalancing
Model cards, decision documentation
Responsible, accountable, consulted, and informed (RACI) assignments, sign-off workflows
Data classification, consent management
Access controls, adversarial testing
Quarterly reviews, regulatory tracking