What is MLOps?

MLOPS CONCEPTS

MLOps(머신러닝 운영)는 DevOps(소프트웨어 개발 및 운영)의 원칙을 머신러닝 모델의 개발, 배포, 운영에 적용한 접근 방식

- 이는 머신러닝 모델을 생산 환경에서 안정적이고 지속적으로 관리하고, 모델의 개발부터 배포 및 운영까지의 전체 생애주기를 체계적으로 지원하는 데 목적이 있음

MLOps의 주요 목적

- 효율적인 머신러닝 모델 개발: 머신러닝 모델을 개발, 실험, 검증하는 과정을 자동화하고 협업을 강화.
- 신속하고 안정적인 배포: 개발된 모델을 빠르고 안정적으로 배포하여 실시간 서비스를 가능하게 함.
- 운영 및 유지관리: 배포된 모델의 성능을 지속적으로 모니터링하고, 성능 저하 시 재학습 및 업데이트를 자동화.

MLOps의 구성 요소

- 모델 개발 (Development):

데이터 준비, 모델 학습, 하이퍼파라미터 튜닝, 실험 관리.
버전 관리 툴을 통해 데이터와 모델 변경 이력을 추적.
도구: Jupyter Notebook, TensorFlow, PyTorch 등.
모델 검증 및 테스트 (Validation & Testing):

- 학습된 모델의 성능을 평가.

데이터 드리프트 및 모델 드리프트를 감지하여 성능 저하를 방지.

- 모델 배포 (Deployment):

모델을 API 또는 애플리케이션으로 배포.
배포 방식: 실시간 배포, 배치 배포, A/B 테스트 배포, Canary 배포 등.

- 모델 운영 및 모니터링 (Operations & Monitoring):

운영 중인 모델의 성능 및 데이터의 변화를 실시간으로 모니터링.
예측 실패율, 응답 속도, 서비스 가용성을 분석.
문제가 발생할 경우 경고를 보내고 필요한 조치를 자동화.

- 자동화 및 재학습 (Automation & Retraining):

새 데이터가 들어오거나 모델 성능이 저하될 경우, 모델을 자동으로 재학습 및 업데이트.

MLOps의 주요 단계
- 데이터 수집 및 준비: 데이터를 수집하고 정제하여 학습에 적합한 형태로 변환.
- 모델 개발 및 학습: 머신러닝 알고리즘을 사용하여 모델을 설계하고 학습.
- 모델 평가: 학습된 모델의 성능을 테스트 데이터로 평가.
- 모델 배포: 학습된 모델을 프로덕션 환경에 배포.
- 모니터링 및 재학습: 배포된 모델의 성능과 입력 데이터를 지속적으로 모니터링하고, 필요시 재학습.

MLOps의 핵심 원칙
- 자동화: 데이터 준비, 모델 학습, 배포, 모니터링 등 반복 작업을 자동화.
- 협업: 데이터 과학자, 엔지니어, 운영팀 간의 원활한 협업을 지원.
- 재현성: 실험 결과와 모델 학습 과정을 재현 가능하게 유지.
- 확장성: 데이터와 모델 규모가 커져도 시스템이 안정적으로 작동.
- 모니터링 및 피드백: 모델의 성능을 실시간으로 모니터링하고 필요한 경우 조정.

MLOps의 장점
- 프로덕션 모델의 신뢰성 향상: 모델 성능 및 데이터 품질을 지속적으로 모니터링하여 예측 실패를 최소화.
- 생산성 증대: 반복적이고 수동적인 작업을 자동화하여 데이터 과학자의 시간 절약.
- 모델 품질 보장: 모델의 재현성과 버전 관리를 통해 안정적인 결과 제공.
- 운영 비용 절감: 효율적인 운영과 자동화를 통해 비용 절감.

MLOps의 도입 사례
- 전자상거래: 추천 시스템 모델을 자동 배포 및 모니터링하여 고객 경험 개선.
- 금융: 사기 탐지 모델을 실시간으로 모니터링하고 업데이트.
- 헬스케어: 환자 데이터 기반 예측 모델을 배포 및 관리하여 의료 효율성 향상.

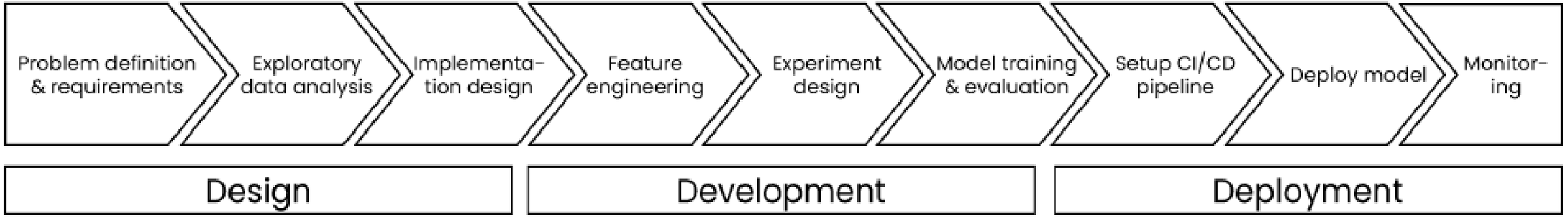
MLOps의 어려움
- 복잡성 증가: 데이터, 모델, 시스템 간의 의존성이 증가.
- 팀 간 협업 부족: 데이터 과학자와 엔지니어 간의 역할 차이로 인해 협업이 어려울 수 있음.
- 운영 환경 관리: 클라우드 및 온프레미스 환경에서의 배포와 관리 복잡성.

결론
- MLOps는 머신러닝 모델을 효과적으로 관리하고 운영하기 위한 필수적인 프레임워크
- 자동화, 협업, 재현성을 통해 머신러닝 모델의 품질과 신뢰성을 높이고, 기업의 데이터 기반 의사 결정을 지원
- MLOps를 성공적으로 구현하면 모델 배포 주기가 단축되고, 운영 효율성이 극대화됨

Machine Learning Operations

*...is the set of practices to **design**, **deploy** and **maintain** machine learning in production **continuously**, **reliably**, and **efficiently**.*

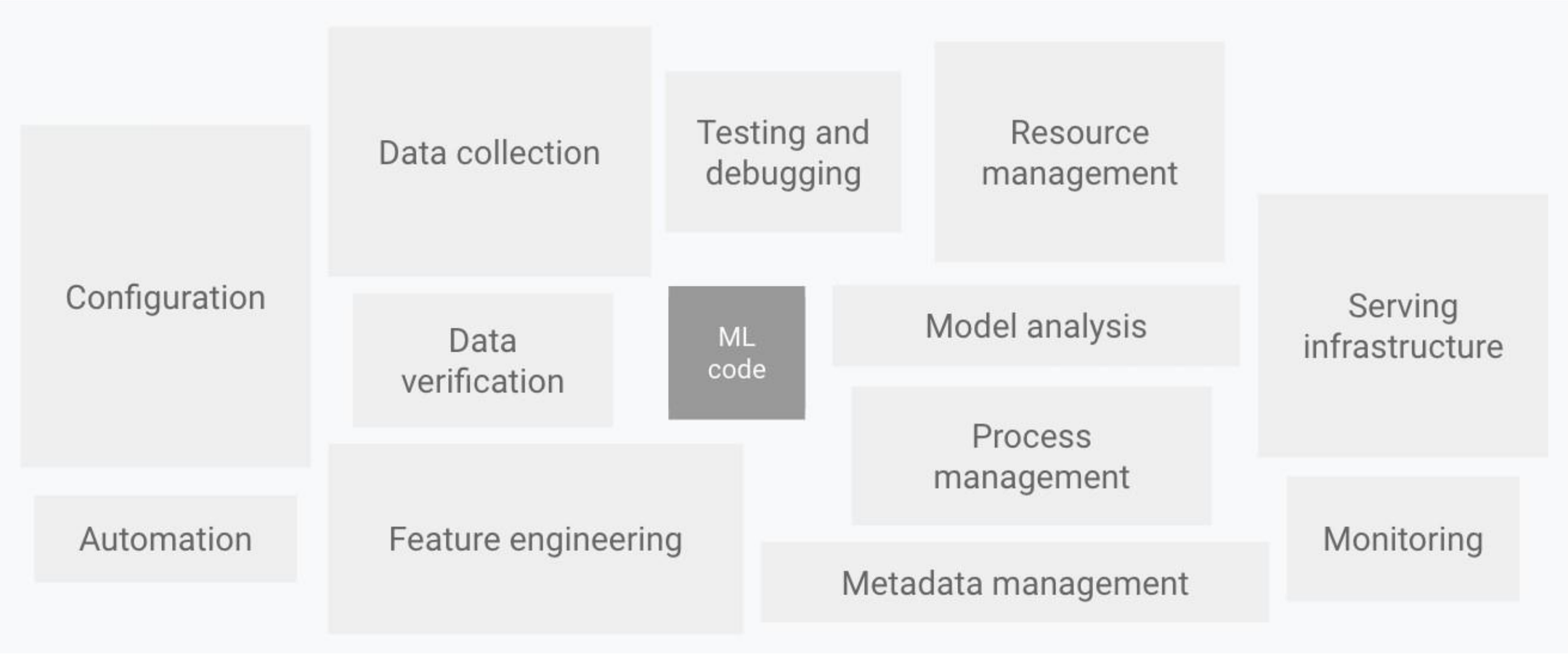
- Focus on machine learning '*in production*'
- The full machine learning lifecycle



The origin of MLOps

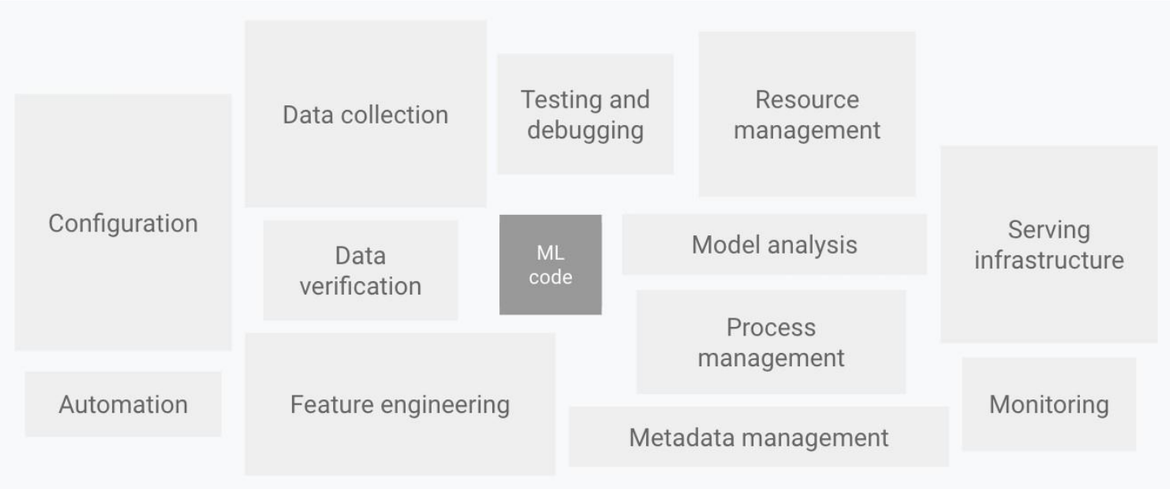
- Practices and tools to deliver software applications
- Development and Operations used to be separate

Why MLOps



[1] https://cloud.google.com/architecture/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning

Why MLOps

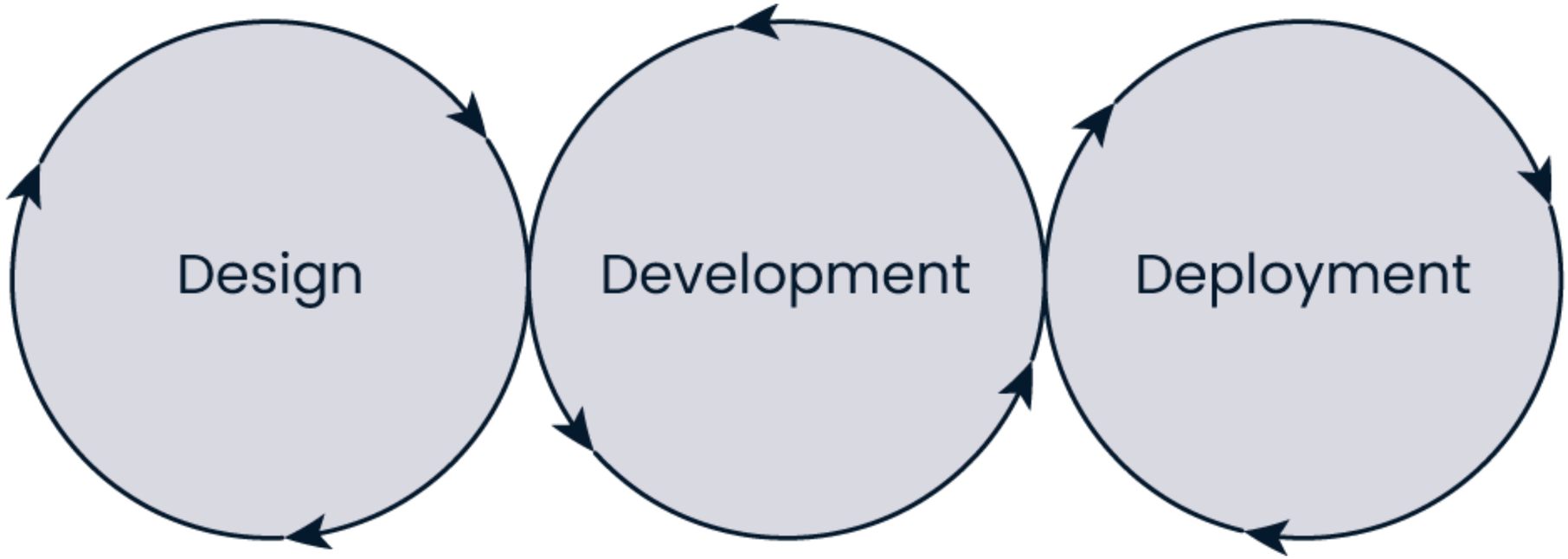


- Improved collaboration
- Automation of deployment
- Monitoring of model performance

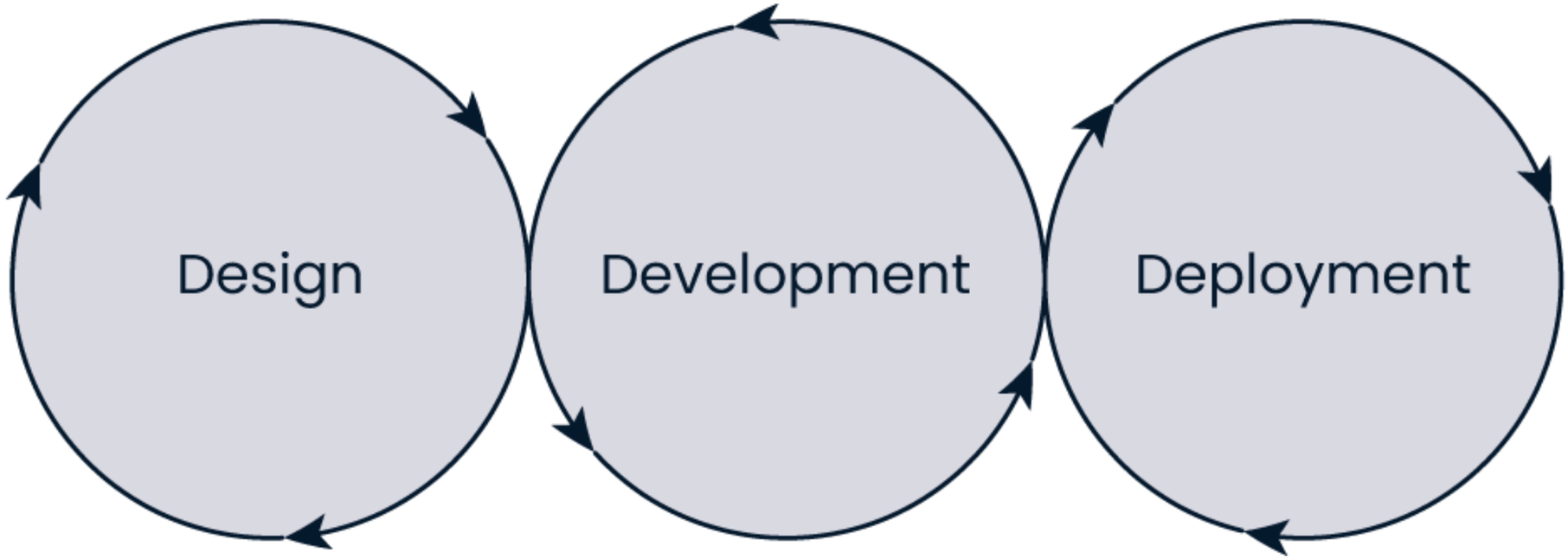
Different phases in MLOps

MLOPS CONCEPTS

MLOps lifecycle

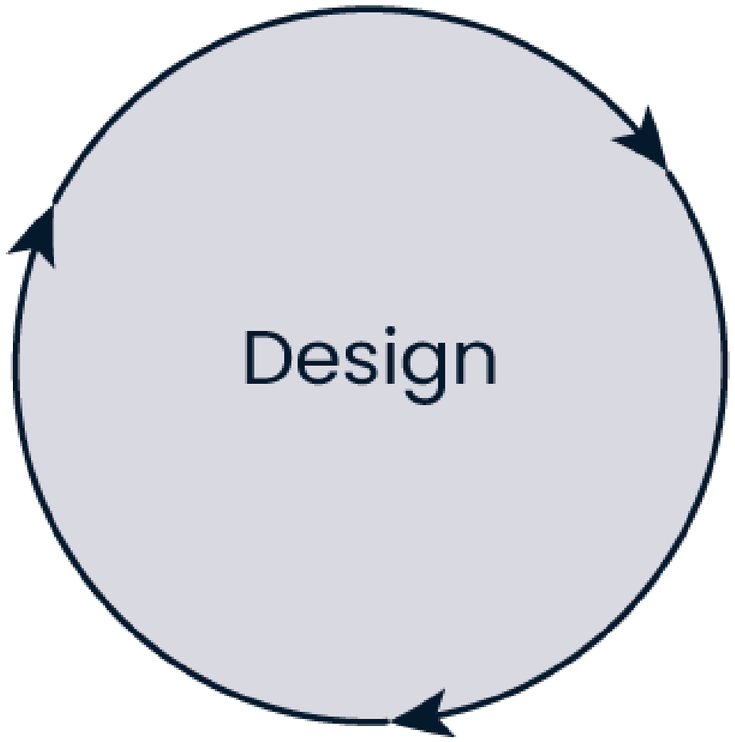


Why the machine learning lifecycle?



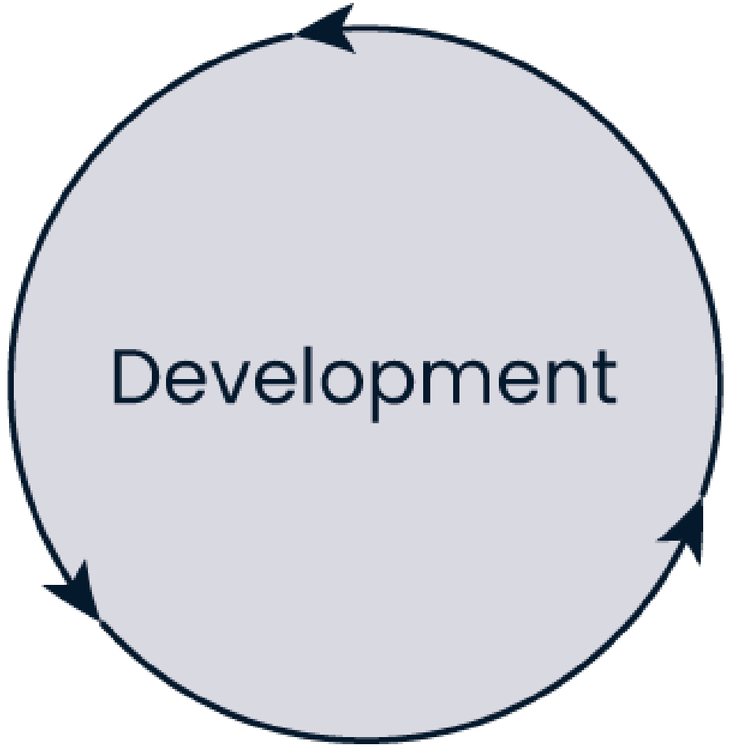
- Structures the ML process
- Defines key players at each stage
- Toolkit for optimization

Design phase



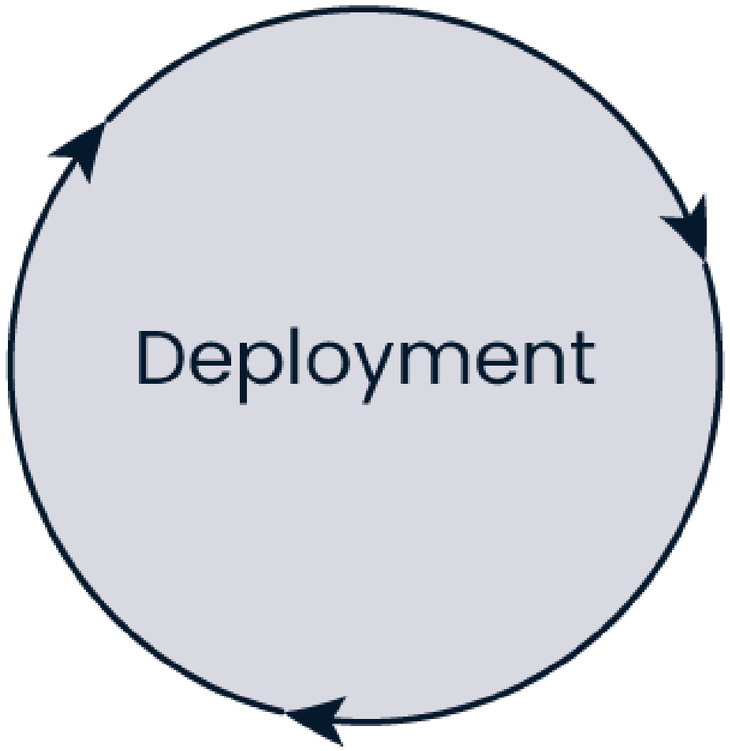
- Context of the problem
- Added value
- Business requirements
- Key metrics
- Data processing

Development phase



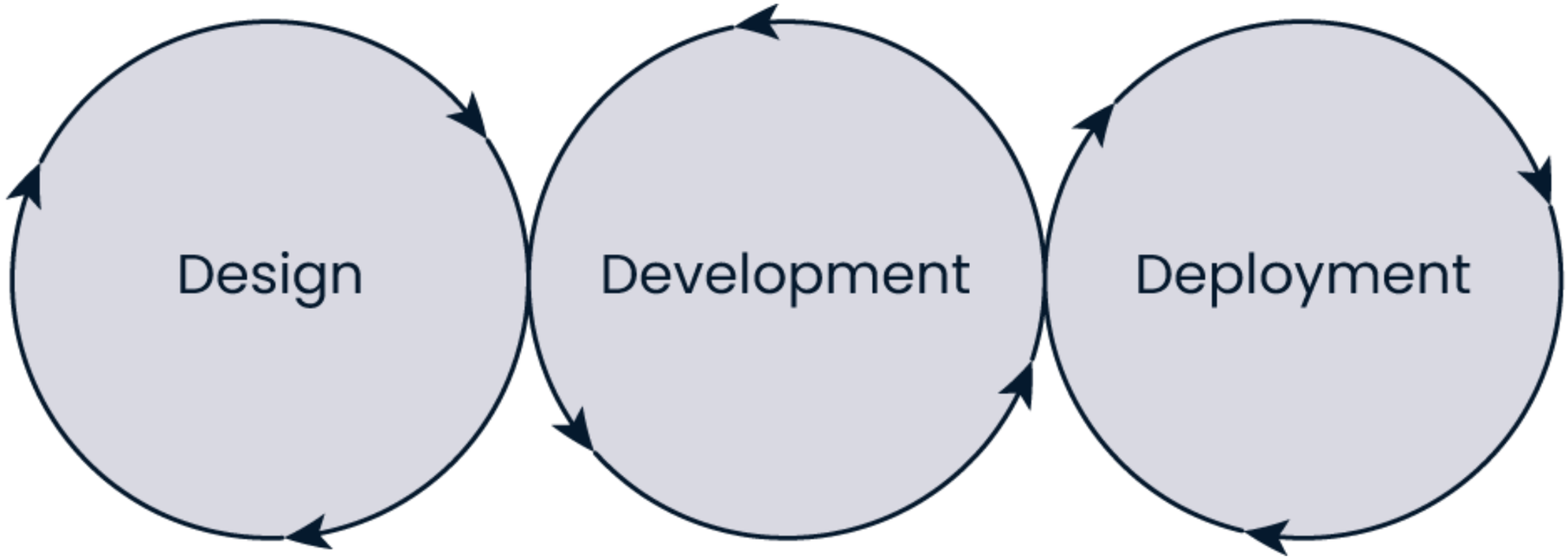
- Develop machine learning model
- Experiment with data, algorithms, and hyperparameters
- Model ready for deployment

Deployment phase



- Integrate the machine learning model in business
- Deploying the model in production
- Monitoring the performance

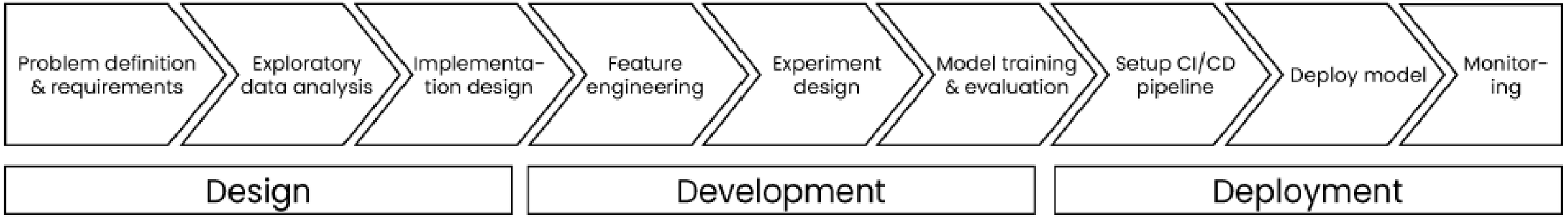
MLOps lifecycle



Roles in MLOps

MLOPS CONCEPTS

Machine learning lifecycle

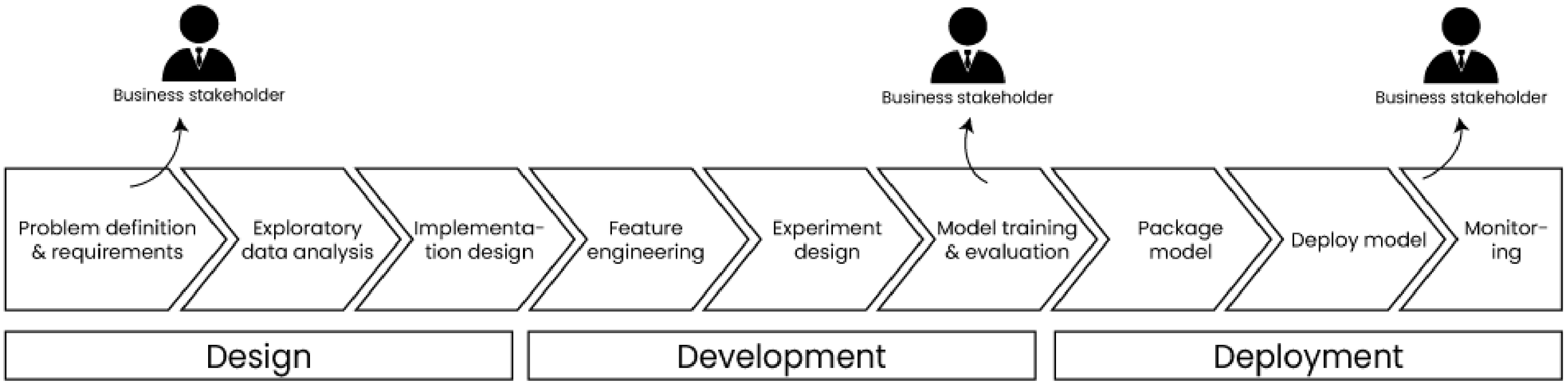


- Business roles
- Technical roles

Business roles

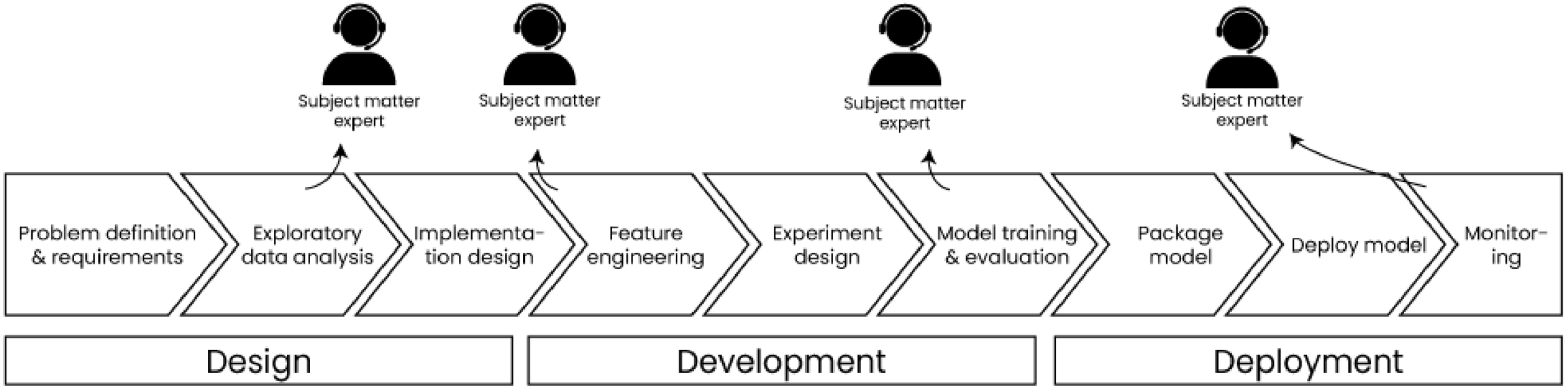
- Business stakeholder
- Subject matter expert

Business roles: business stakeholder



- Budget decisions
- Vision of company
- Involved throughout the lifecycle

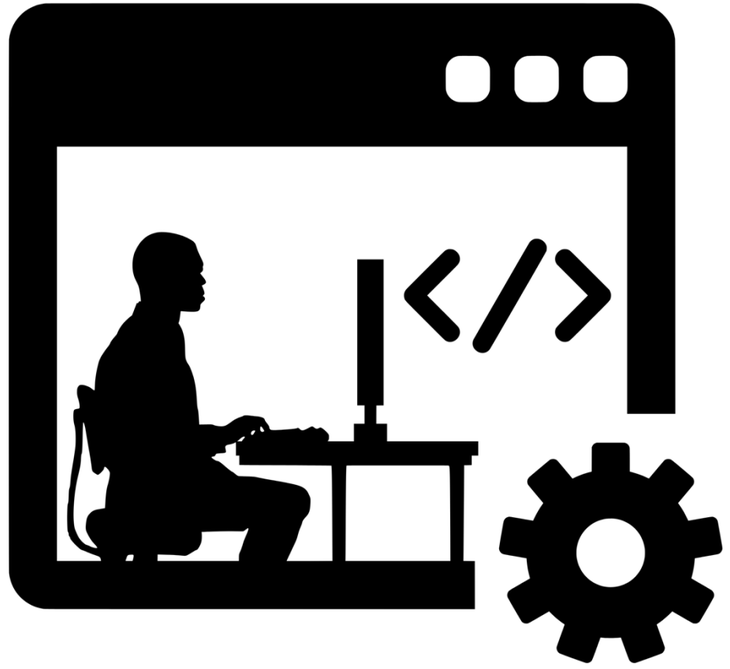
Business roles: subject matter expert



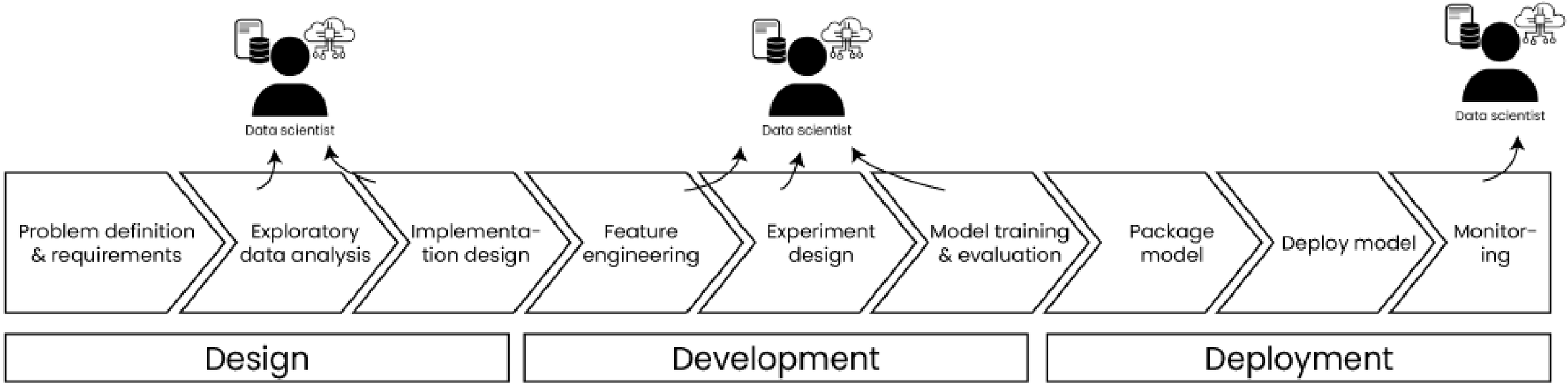
- Domain knowledge
- Involved throughout the lifecycle
- Interpret and validate data

Technical roles

- Data scientist
- Data engineer
- ML engineer

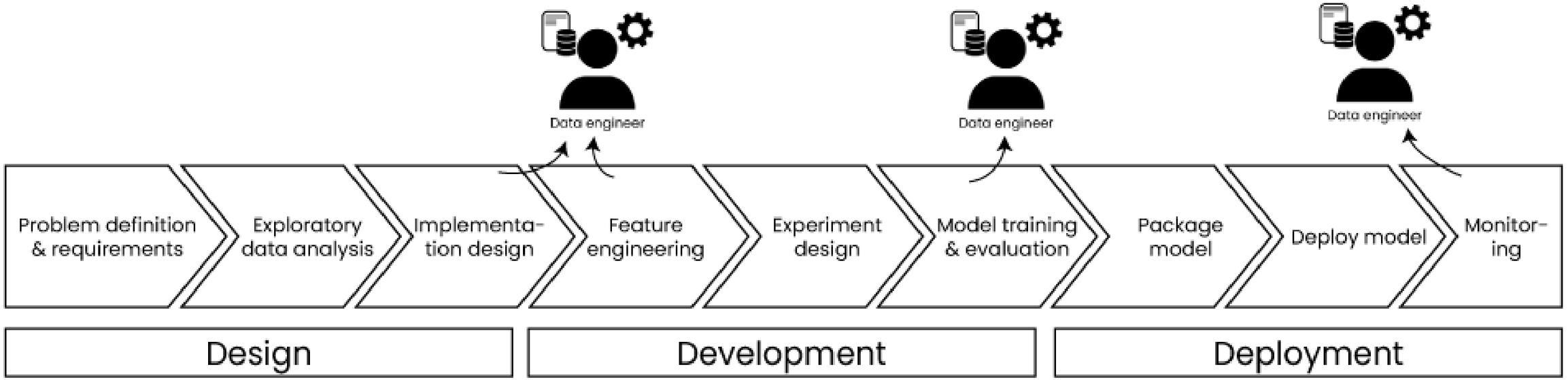


Technical roles: data scientist



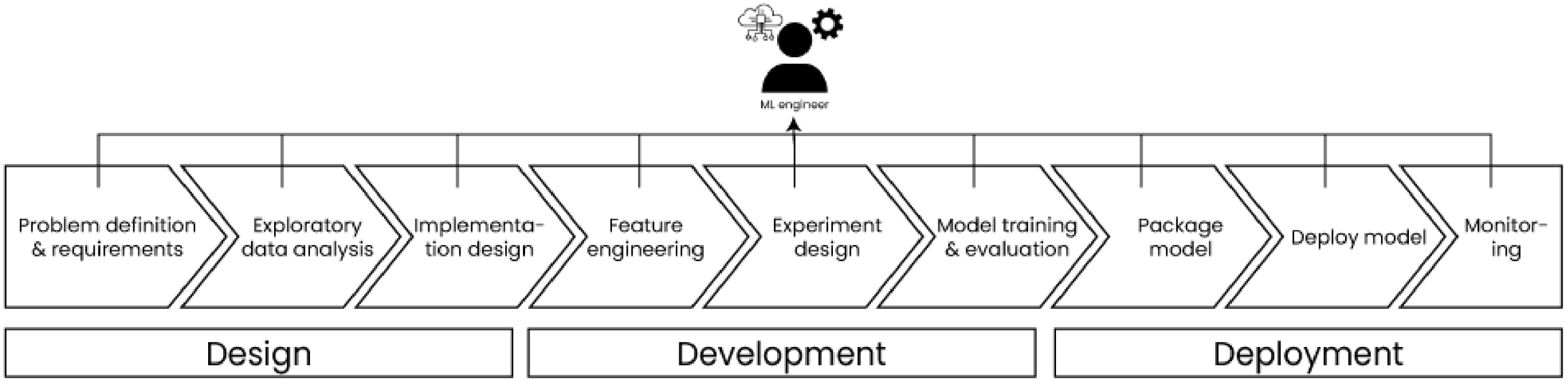
- Data analysis
- Model training and evaluation

Technical roles: data engineer



- Collecting, storing, and processing data
- Check and maintain data quality

Technical roles: ML engineer



- Versatile role
- Specifically designed for complete machine learning lifecycle

Additional roles involved in ML

- Data analyst, developer, software engineer, backend engineer
- Responsbility of roles can vary depending on application of machine learning
- Startup is different from a large enterprise