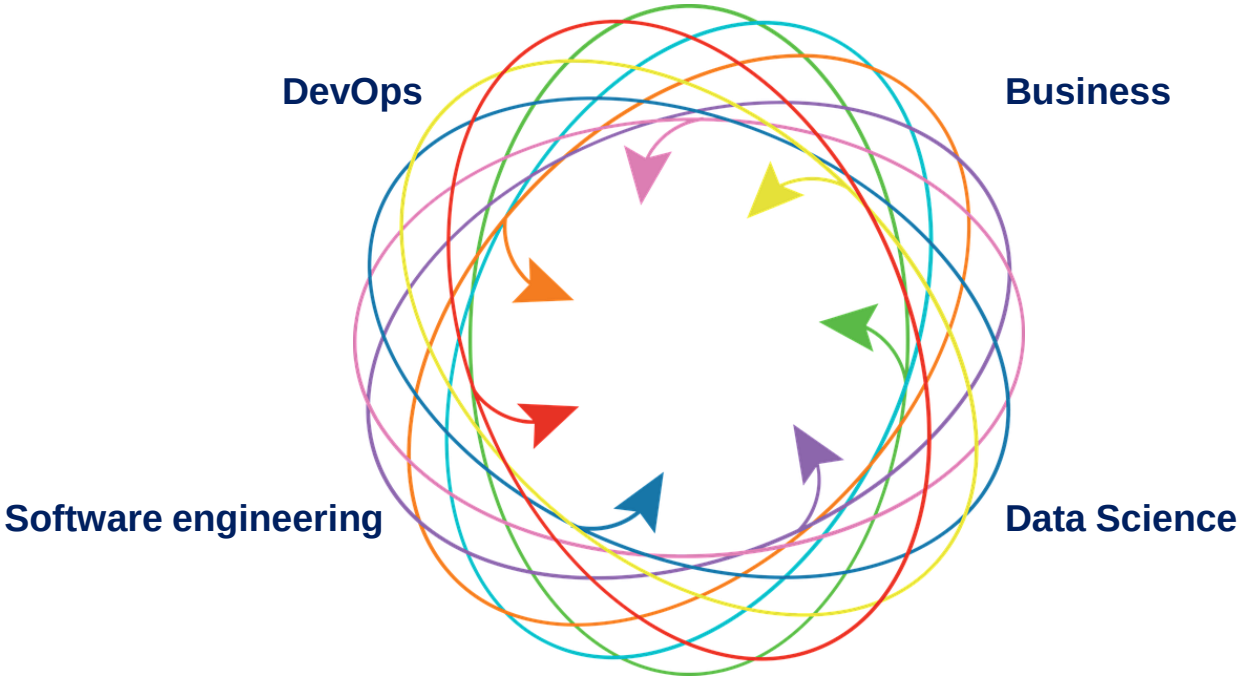
Challenges and Key Principles for ML Systems

Specific Challenges of ML Systems

- The need for reproducibility (versioning everywhere)
- Entanglement
- Data dependencies
- Configuration issues
- Data and feature preparation
- Model errors can be hard to detect with traditional tests
- Separation of expertise

ML System Contributors



The architecture of a production machine learning system needs to take into account business requirements, as well as the unique challenges at the intersection of data science, software engineering and devops.

Research vs. Production Environments

	Research	Production
Separate from customer facing software	✓	x
Reproducibility matters	Sometimes	Almost always
Scaling challenges	x	✓
Can be taken offline	✓	x
Infrastructure planning required	Sometimes	Almost always
Difficult to run experiments	x	✓

Key Principles for ML System Architecture

- **Reproducibility**: Have the ability to replicate a given ML prediction
- **Automation:** Retrain, update and deploy models as part of an automated pipeline
- **Extensibility:** Have the ability to easily add and update models.
- **Modularity:** Preprocessing/feature engineering code used in training should be organized into clear pipelines
- **Scalability:** Ability to serve model predictions to large numbers of customers (within time constraints)
- **Testing**: Test variation between model versions