AWS re:Invent

AIM202-S

Artificial intelligence in Healthcare

Matthew Rich
Principal
PwC

Sidd Bhattacharya
Healthcare AI
PwC

JR Burch
Manager, Information & Digital Solutions
Eli Lilly

AWS re:Invent

AI at PwC

We infuse and accelerate innovation into the business

Dream it.

Strategy and Research

Our dedicated research specialists and subject matter experts continuously track a knowledge base of 265+ emerging technologies and share our perspective on how these technologies impact the marketplace.

- Immersion sessions
- Research dossiers

Build it.

The Lab

Our lab puts research to the test through hands-on experimentation, rapid prototyping, and proofs of concept (POCs) that demonstrate the business potential of emerging technologies. We execute 30–40 POCs annually.

- Client sessions in lab
- Prototype co-creation
- Demonstrations

Live it.

Sales and Delivery

The digital solutions team includes delivery consultants and specialized teams that deliver new services and digital solutions, at speed and at scale.

- Pre-built, flexible data and analytics solutions

AI areas of focus

Machine Learning
Using algorithms to learn from data and solve business problems without being explicitly programmed

Deep Learning
Leveraging cutting-edge machine learning algorithms inspired by artificial neural networks especially for unstructured data

Natural Language
Understanding human speech and text through application of computer science, AI, and computational linguistics

Data Engineering/Model Ops
Using cutting-edge architecture to analyze terabytes of data and deploy AI models on cloud for production

Simulation & RL
Testing various scenarios in models of real-world processes and finding optimal strategies under those scenarios

Automated ML
Automating and standardizing machine learning pipelines to make them more accessible and reproducible

Embodied AI
AI that is embodied within a physical artifact and interacts with humans and the environment (robots, IoT, autonomous drones)

Responsible AI
Developing fair, safe, explainable, accountable, and ethical AI with a combination of people, process, technology, and governance

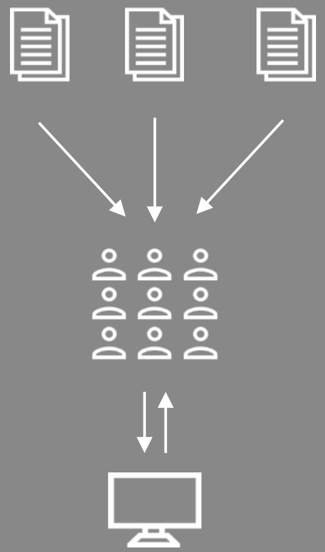
Use cases of AI in Healthcare

AI for intelligent case management in drug safety, aka Pharmacovigilance (PV)

AI shifts the experience from a highly manual & labor-intensive process…

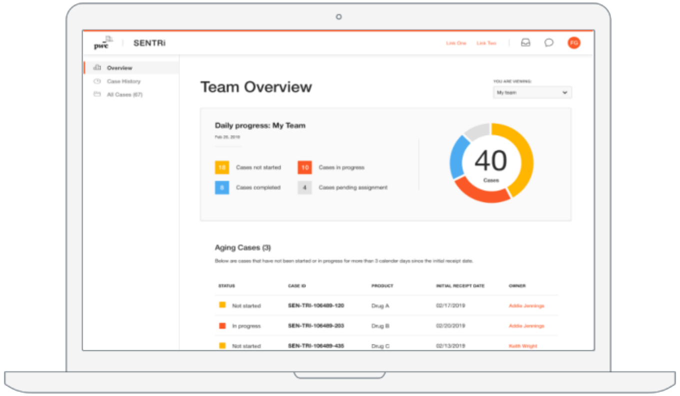
…to one that is highly intelligent, dynamically learning, and automated

Data is manually aggregated from multiple sources in a multitude of formats



Dozens of manual steps are required to intake, enter, and interpret the data just to determine case seriousness; only then can the real decision process begin

Machine learning and NLP optimize processing by automating data intake and entry, extracting key information, and interpreting that data



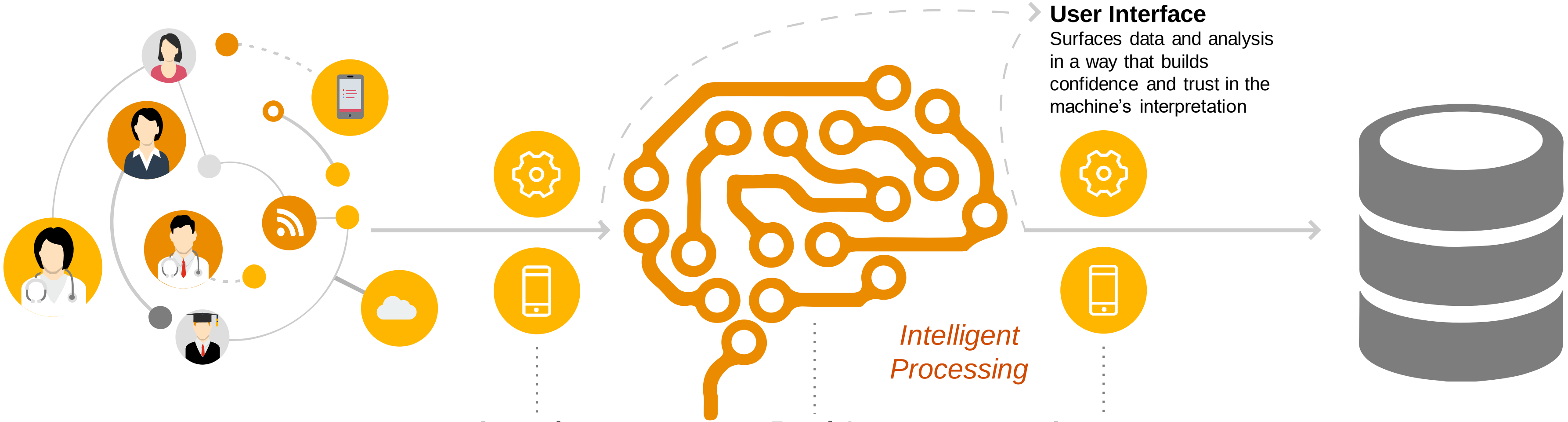
Intelligent automation (IA) reduces the effort the team spends on these transactional tasks, enabling them to focus on other activities, improving operational efficiency

The solution uses intelligent automation to optimize case processing and is based on user-centric design

Various Digital Sources

Intake, Triage, Initial Processing, and Transfer to Safety Database

Safety Database



Including PSPs, EDC/digital sources, and classification of reports (e.g., product complaints, adverse events, other inquiries)

Ingestion
Cloud-hosted data lake to ingest data from multiple sources

Read & Recognize
Pre-trained NLP models to identify key entities

Interpret
Machine learning and rules-based models to perform interpretation

Case sent to the right step in the workflow depending on completeness and risk identified by the interpretation and verified by the user

The solution has increased case processing efficiency by 30% as a result of intelligent automation

75%

reduction in duration of processing

>30%

reduction in overall effort to process a case

Issue & event detection using AI at call centers

Automatically detect reportable issues & events in call center data

Issue: A large pharmaceutical company, receiving a very high call volume through its customer engagement programs, is facing ***challenges in issue detection***. Failure to identify reportable issues (e.g., missed dosage, off label usage) causes ***inefficiencies in terms of cost, effort, and compliance***. Large batches of incoming reports with a few missed issues and events (if found during regular quality checks) need to be completely reviewed and evaluated again to stay compliant.

Solution: ***A combination of rules and ML models can be developed to augment the manual effort required for accurately and quickly identifying reportable issues*** in call center reports, thus significantly reducing inefficiencies/costs associated with missed reportable issues.

We conducted a POC to demonstrate that select rules can be applied to identify repeatable events and issues to aid call center agents.

POC Approach

Speech to Text Transcription → NLP Processing → Named Entity Recognition (NER) → Drug, Dosage, Frequency → Adverse Events

Step 1: Used Amazon Transcribe to convert voice recordings from call center to text

Step 2: Used AI/ML tools to teach the engine how to identify drugs, dosage, frequency, and AEs

Step 3: Developed an engine to:

A. Extract drugs & AEs
B. Compare extracted drug info with drug guidelines
C. Issue warnings for off label dosage, missed dosage; AEs identified to alert agents

Next step is to conduct a pilot with the client using the POC engine to demonstrate improvement in business process.

Other Healthcare use cases

Medical document retrieval chatbot

Retrieve documents using a chatbot that understands natural language queries

What's the challenge?

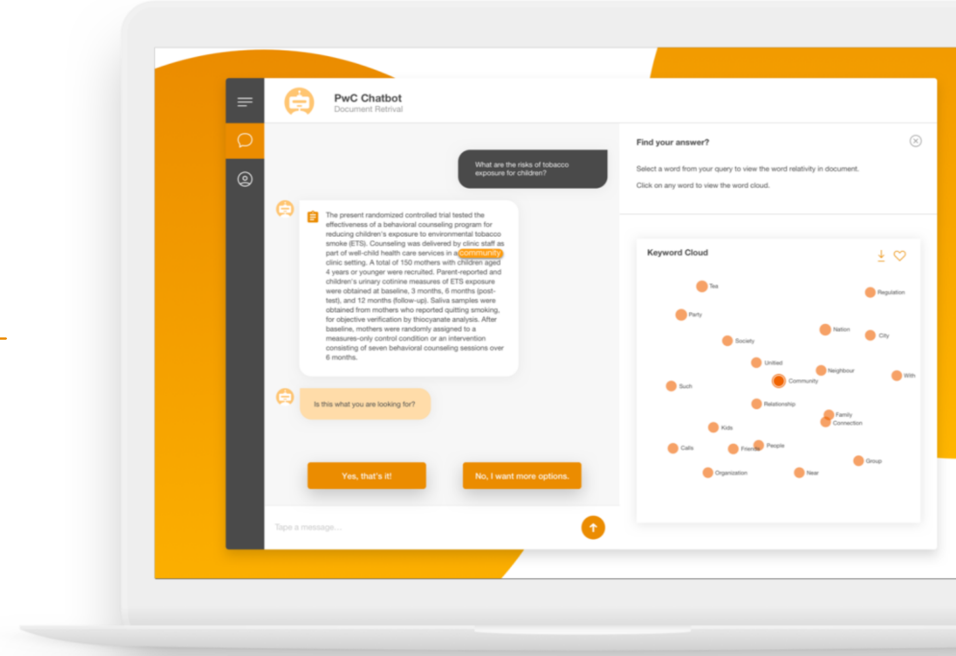
Traditional document retrieval tools rely on handwritten rules, word matching, and metadata in order to retrieve documents that are relevant to a user's query. The main challenge is that traditional models cannot understand natural language since it would be impossible to handwrite all rules of language.

Our solution

Our solution is to train a deep learning model to understand the natural language of user queries and match them to the natural language of each document in a database. This model was trained using a dataset of over 14 million medical document abstract-title pairs.

What's the business value?

Our solution enables us to build a chatbot for any domain by training a deep learning model on a dataset of document abstract-title pairs. By deploying an early-stage chatbot, data can also be collected through usage, which would provide more task-specific training examples for edge cases.



Regulatory policy evaluation

Identify requirements in regulations and gaps in existing policies using natural language processing (NLP) algorithms

What's the challenge?

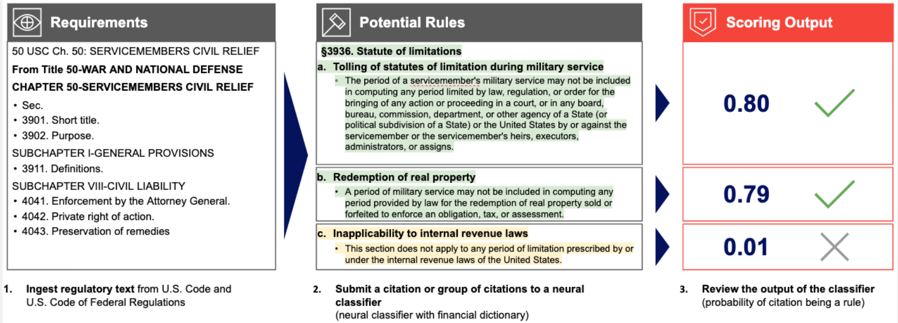
Complying with regulatory obligations that are global, changing, and difficult to monitor is repetitive and manually intensive. Existing platforms and technologies use brute-force techniques and provide limited technology relief (pencils and notepads).

Our solution

Automatically identify rules from regulatory texts such as the US Code (USC), Code of Federal Regulations (CFR), and supervisory letters using NLP. Match and rank rules against policies to identify rules requiring further investigation using unsupervised NLP.

What's the business value?

Rules are correctly identified 75% of the time in the external law and/or regulation (Step 1), and policy review time can be reduced by ~50+% (Step 2).



Unstructured – Information extraction

Machine learning models that extract information from PDF and Excel documents

What's the challenge?

Many business processes (e.g., reviewing contracts) rely on extracting information from unstructured documents such as PDF, Word, and Excel, which can be error-prone and time-consuming.

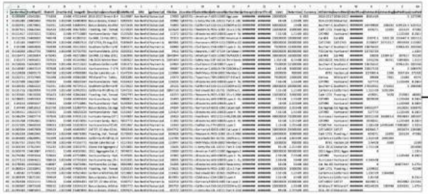
Our solution

We developed machine learning models that can read through unstructured documents such as PDF, Word, and Excel and can extract a variety of fields (e.g., dates, currency amounts) and prepare model outputs for downstream consumption in other business processes.

What's the business value?

With a target accuracy of 90+%, clients can extract the majority of fields present in unstructured documents, reducing the time spent on the process by thousands of hours and improving data capture.

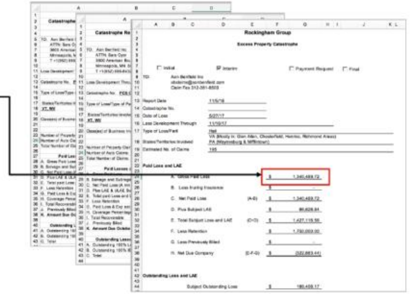
Metadata (i.e. annotations, fields & values, ground-truth, golden data, etc.)



Search for the ground-truth data in the raw claims documents

Calculate the the hit-rate (i.e. count of values found over count of values provided)

Raw underlying claims documents



Deeper dive

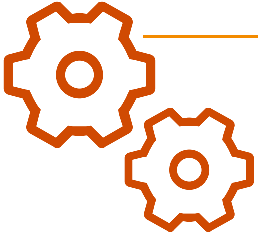
A closer look at the technology behind the intelligent automation in drug safety

The current release of the platform deploys an end-to-end case narrative processing pipeline

Example Text

Patient suffered from diarrhea once this month and was later admitted to the hospital complaining of severe stomach cramps.

NLP Engine



1 Medical Text Classification

Patient History

2 Named Entity Recognition (NER)

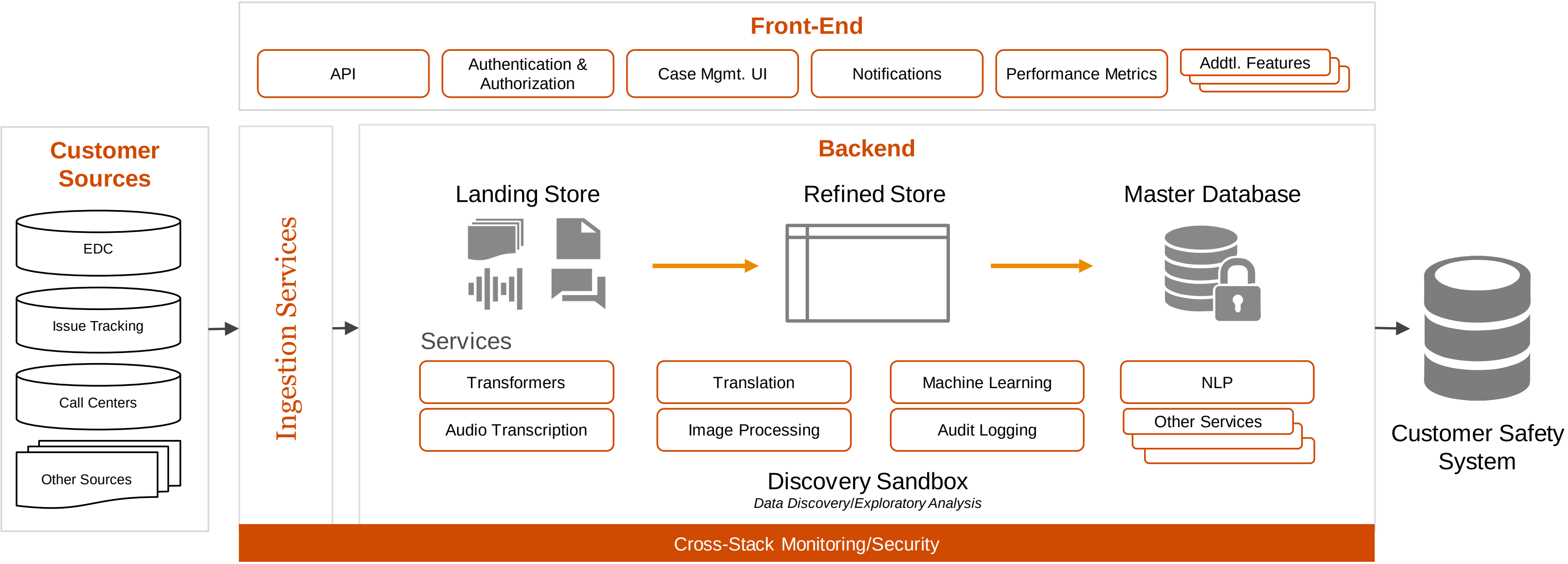
	Entity	Type
0	diarrhea	event
1	stomach	location
2	stomach cramps	event

3 Term Similarity Search

As Reported term	Most Similar Term
diarrhea	Diarrhoea (excl infective)
stomach cramps	Gastrointestinal and abdominal pains (excl oral and throat)

The platform is built on a cloud-native, serverless architecture to enable scaling on demand

The system follows an event-driven serverless architecture leveraging HIPAA-compliant cloud-native technologies to provide the necessary capability to scale on demand



Contact us

Matthew Rich
Partner
matthew.rich@pwc.com

JR Burch
Manager, Information & Digital Solutions, Eli Lilly
burch_jacob_r@lilly.com

Sidd Bhattacharya
Director
siddhartha.bhattacharya@pwc.com

Additional topics

End-to-end machine learning model lifecycle

Process and tools

End-to-end modeling process involves using a variety of disparate tools and technologies

Multiple handoffs between data engineers, data scientists, and DevOps engineers

Data scientists need to iterate with agility on building machine learning models, while application engineers need to integrate the model changes in a repetitive and automated manner into stable production systems

Modeling Process

Raw Data Collection Batch & Streaming

- Data collection and storage in distributed data stores
- Streaming data ingestion pipelines

- Kafka
- Amazon S3
- Hadoop
- Cassandra

Data Engineering

- Data cleansing, preparation, metadata management, data partitioning, ETL pipelines

- Spark
- Hive
- Airflow
- Presto
- Apache Zeppelin

Model Training

- Labeling, feature engineering, model selection, hyperparameter tuning

- TensorFlow
- PyTorch
- Scikit-learn
- Horovod
- Keras
- Jupyter

Experiment Tracking/ Data Provenance

- Model, code, data, and config tracking

- Pachyderm
- DVC
- Weights & Biases
- MLflow

Model Deployment/ Monitoring

- Model serving at scale for inference

- Docker
- Kubernetes
- Prometheus
- Clipper

Process Enablers

- Cloud Platforms
- Container Orchestrators
- Distributed ML Frameworks
- Continuous Integration
- Continuous Deployment
- Monitoring

Real-time prediction model deployment challenges

Building and deploying machine learning systems involves a very different set of architectural practices compared to traditional software engineering.

Organizations are building more machine learning models, but only a few make it to production.

Complexity

- Wide variety of languages are used to train models (e.g., Python, R , Scala, Java)
- High-velocity open-source ML frameworks and different ways to serve models (e.g., TensorFlow, Keras, Scikit-learn, PyTorch, Spark ML)
- Library versions and "dependency hell" cause difficulties in transferring models from development to production environments
- Different technology stacks are used for training models versus deploying models

Lack of Standard Methodology

- Custom tooling developed in-house in the absence of established methods (e.g., FBLearner – Facebook, TFX – Google, and Michelangelo – Uber)
- Fragmented tools and standards that focus on a very specific problem but do not integrate into the larger end-to-end process

Model Diagnostics

- Monitoring model latency, tracing API requests, and collecting request metrics such as request rate & error rate involves wiring together several custom tools

Maintenance and Scalability

- Machine learning models are compute intensive; extensive hardware resources are required to scale machine learning models to serve predictions at large scale
- Rolling out new versions of a model and rolling back deployments can be time-consuming, effort-intensive, and error-prone
- Infrastructure maintenance and upgrades can cause downtimes in the model operations
- Infrastructure needs to be elastic to scale up or scale down based on demand

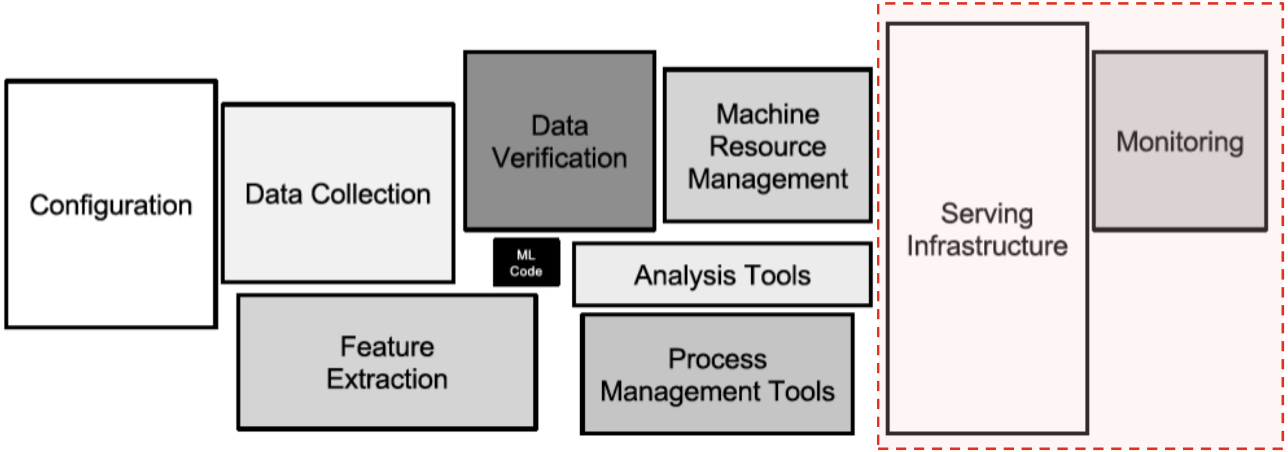
Concept Drift and Feedback Mechanism

- Model predictive performance can change over time due to non-stationary distributions of real-world data
- Monitoring predictive accuracy of models in real time and comparing to ground truth is difficult in most use cases
- Monitoring model metrics over time and triggering retraining of models via a feedback mechanism is a non-trivial process

Automation

- Simple, fast, and repeatable pipelines to deploy incremental changes from development to production environments
- Automation of model build and integration processes

“Hidden Technical Debt in Machine Learning Systems”

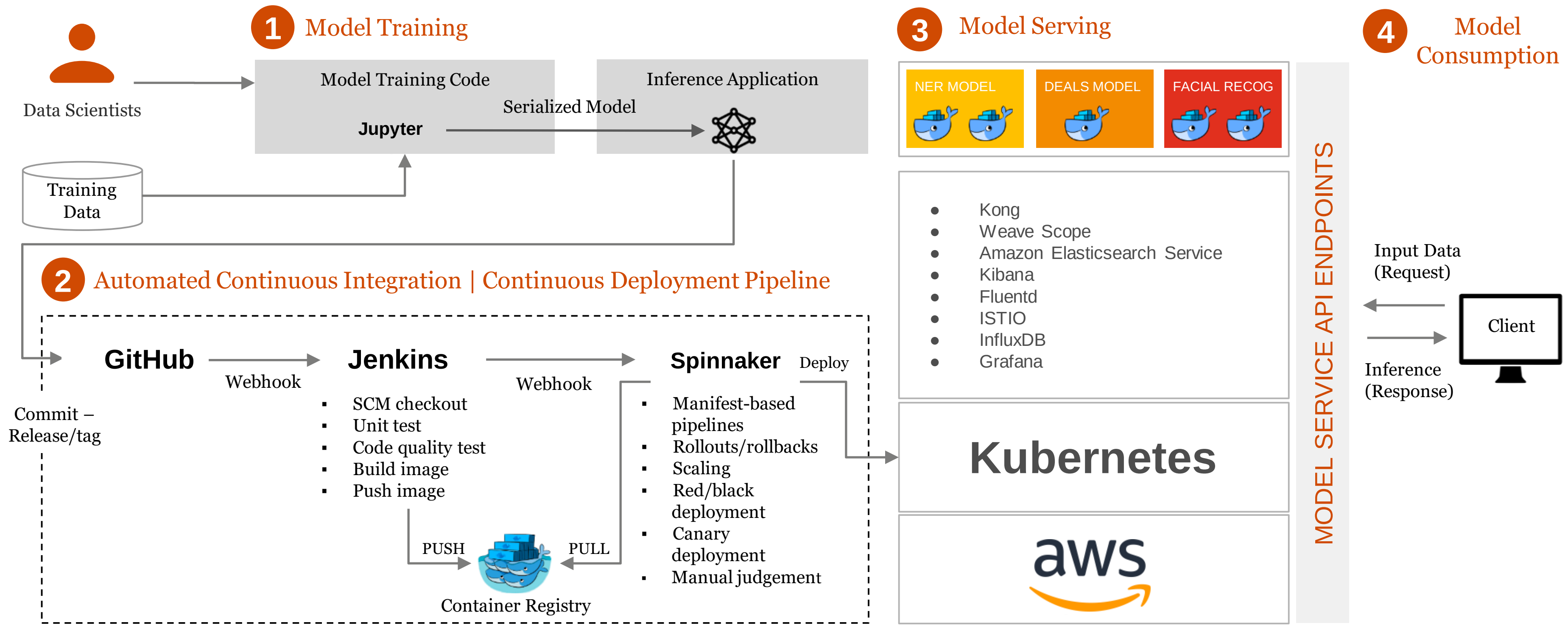


Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

Paper by:
D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, M. Young, J.F. Crespo, and D. Dennison. Hidden Technical Debt in Machine Learning Systems. In Neural Information Processing Systems (NIPS). 2015.

Real-time model operations platform: Modelhive

High-level process architecture & technology stack



Contact details

Thank you

pwc.com

Please complete the session survey in the mobile app.