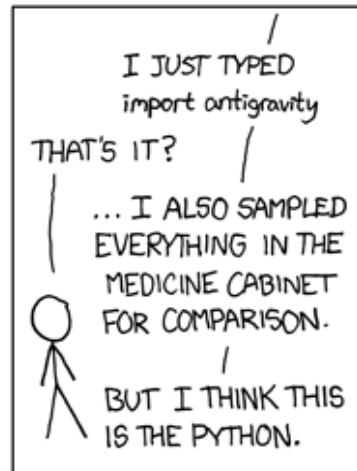
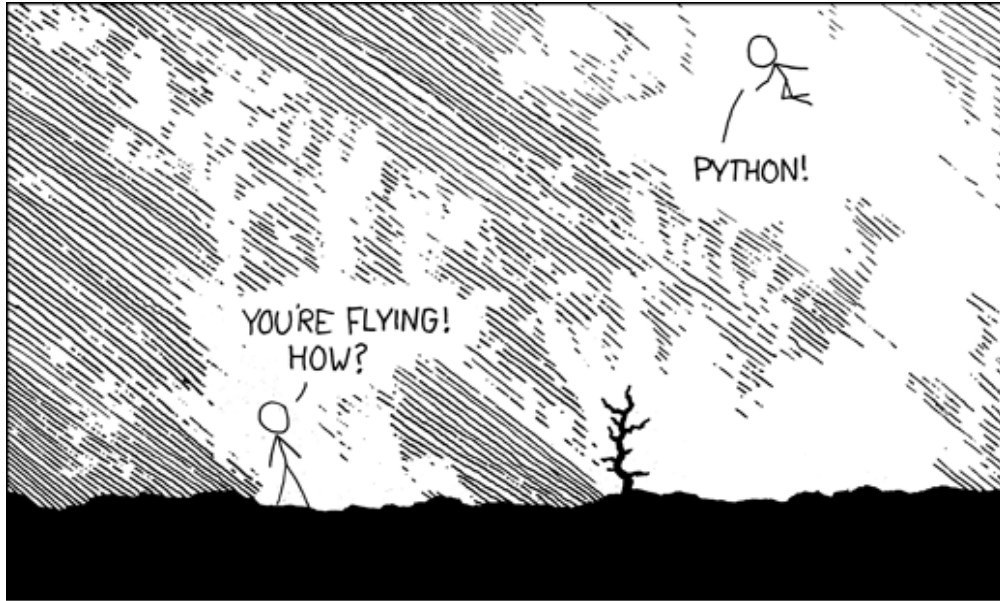


# Demonstrate latest developments in Python to the application of machine learning to business related problems

Tjun Hoh

# Why Python?

# Python can do anything



# Python pros and cons

- Semicolon-less
- Bootcamp-able, Medium-able
- ChatGPT-able, Github Copilot-able
- Libraries, libraries, libraries (have I already mentioned libraries?)
- APIs
- SDKs
- Cons: none

# Python is in demand

## Top Ten Specialized Skills in 2022 AI Job Postings in the United States, 2010–12 Vs. 2022

Source: Lightcast, 2022 | Chart: 2023 AI Index Report

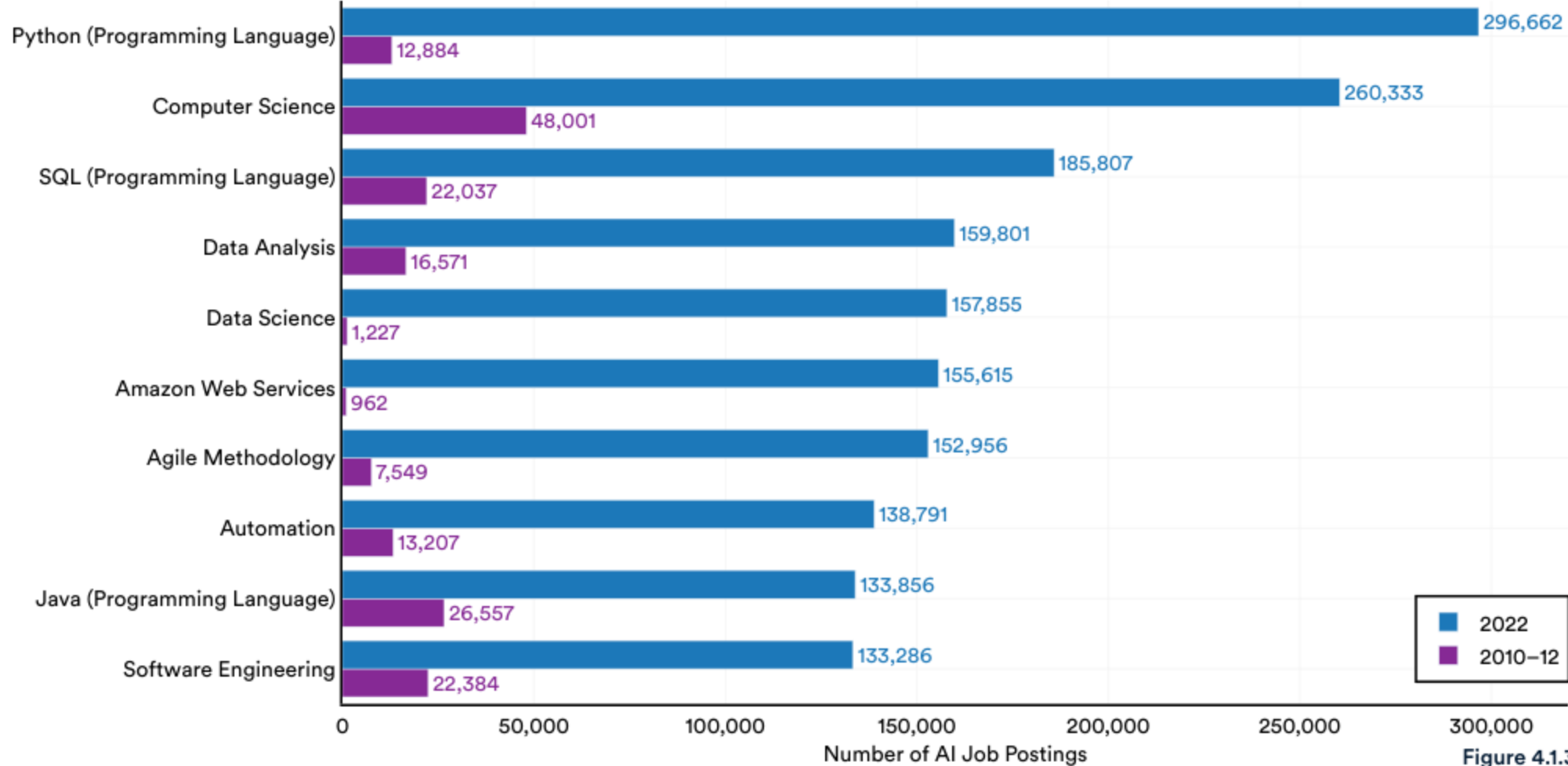


Figure 4.1.3

# ChatGPT can write Python



**Anonymous**

I have a dataset detailing customer purchase habits, can you write python code for a recommender system please?



**ChatGPT**

Absolutely, I can help you create a basic recommender system. There are various types of recommender systems, but a common and relatively straightforward type is the collaborative filtering recommender system.

*ChatGPT*

# Python milestones

- Python 3.0
  - `print("Hello world!")` #A miracle!
- pandas / NumPy / SciPy / matplotlib
- scikit-learn / PyTorch / TensorFlow / Keras / LangChain
- Libraries, libraries, libraries
- Anaconda / Jupyter

**Machine learning?**



# Statistics, ML, and AI

- Classical statistics and machine learning
  - regression, classification, clustering
- Deep learning:
  - neural networks, autoencoders, transformers
- Software paradigms:
  - low-level libraries, high-level APIs
- Hardware paradigms:
  - GPU computing, quantum computing

# Large language models

- ChatGPT
- Github Copilot
- Google Gemini
- Llama

# Communities / repositories

- HuggingFace
- Kaggle
- There's An AI For That
- arXiv

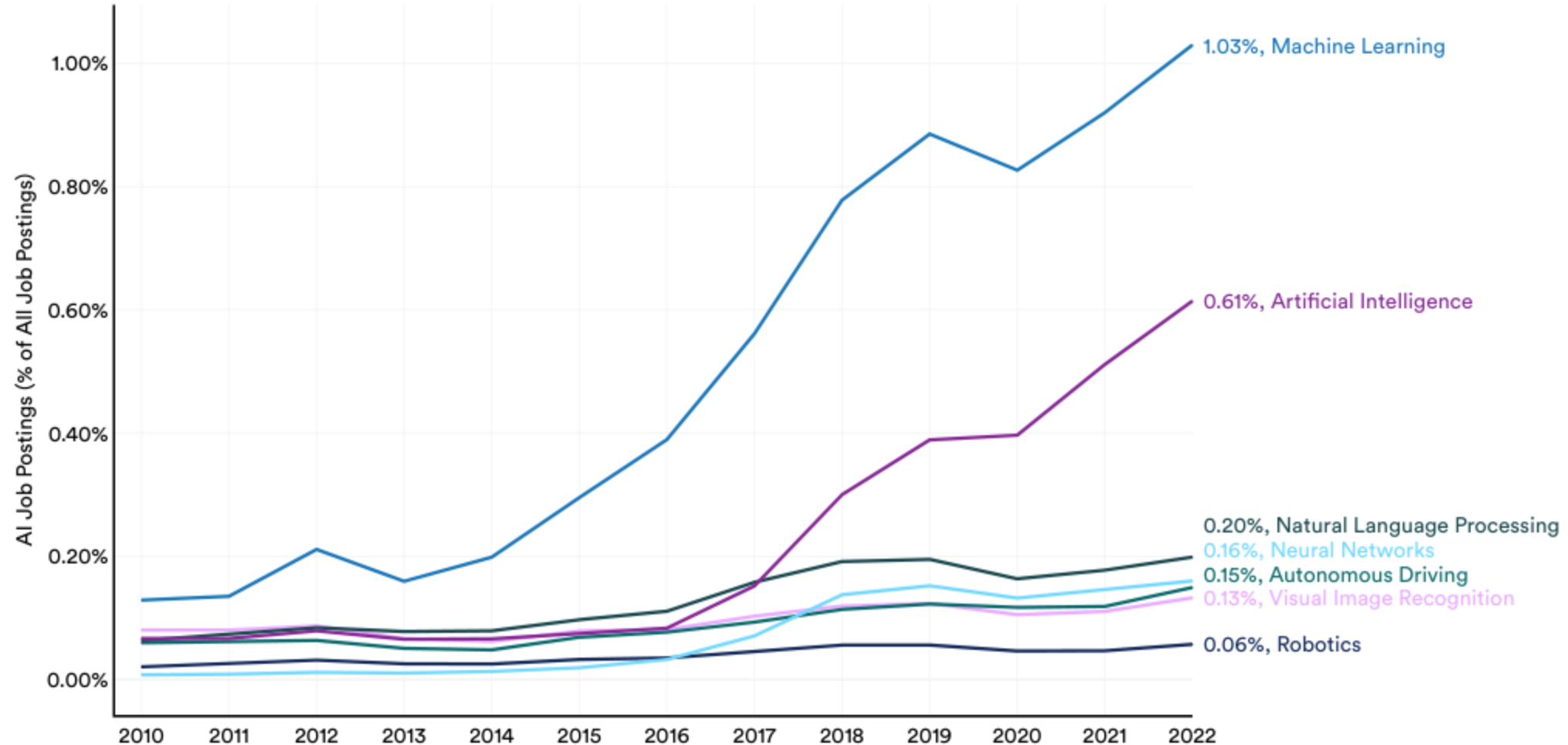
# ML pipeline

- End-to-end:
  - Pandas → NumPy → scikit-learn → PyTorch / TensorFlow
- Automation (AutoML):
  - PyCaret, auto-sklearn, AutoKeras
- Interfacing with infrastructure:
  - Libraries, APIs, SDKs
  - SQLAlchemy, openai, AWS SDK for Python, Faust

# Machine learning is in demand

## AI Job Postings (% of All Job Postings) in the United States by Skill Cluster, 2010–22

Source: Lightcast, 2022 | Chart: 2023 AI Index Report



**Are we in business yet?**

# Business operations

- Process analysis
- Waiting time analysis
- Supply chain management
- Quality management
- Demand forecasting
- Revenue management

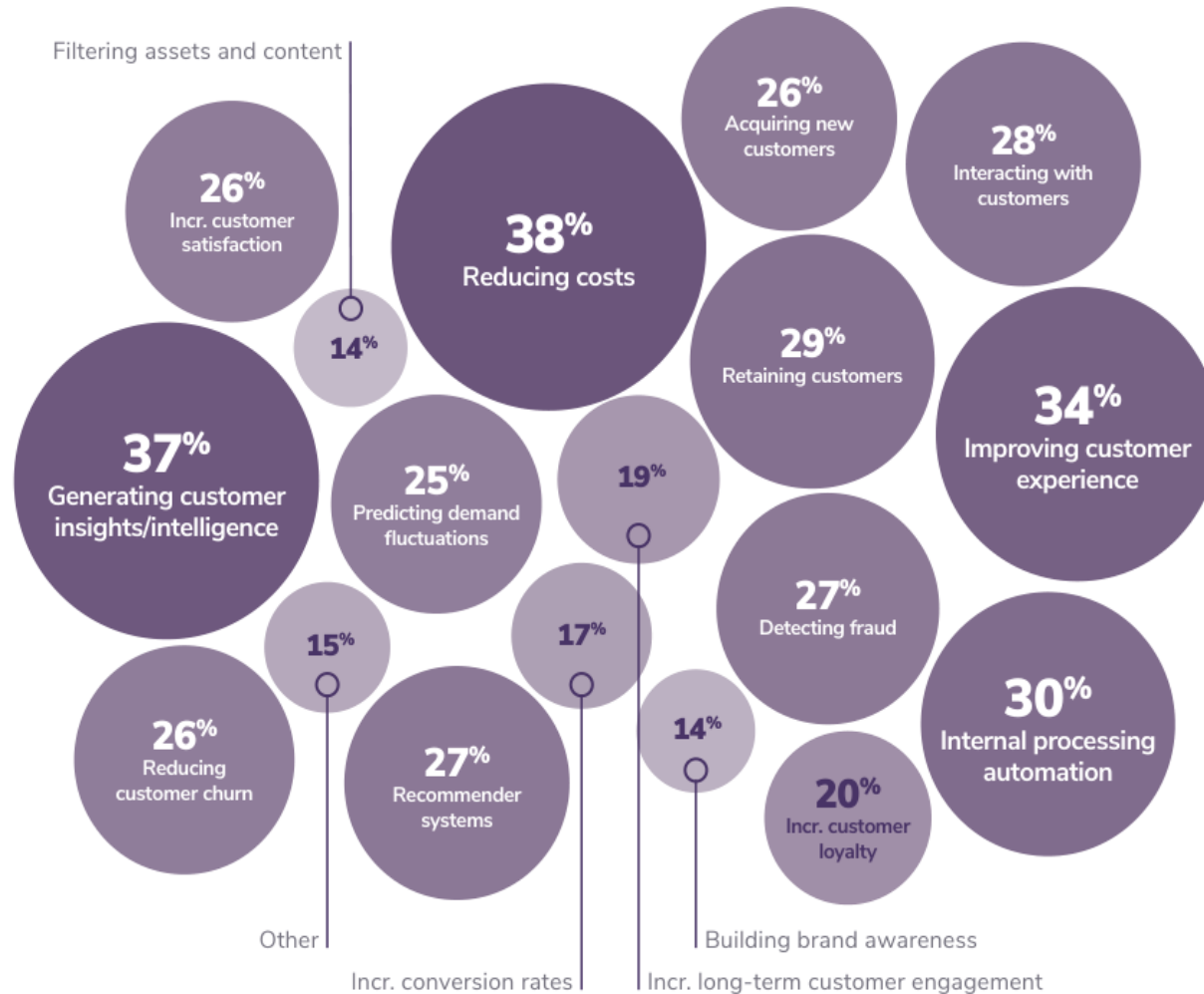
# ML use-cases

- Predictive analytics / forecasting
- Recommender systems / clustering / NLP
- Decision theory / strategy
- Data visualisation
- Operations optimisation
- Network analysis
- Image processing



# ML use-case frequency

## Machine learning use case frequency



*2020 state of enterprise machine learning, Algorithmia*

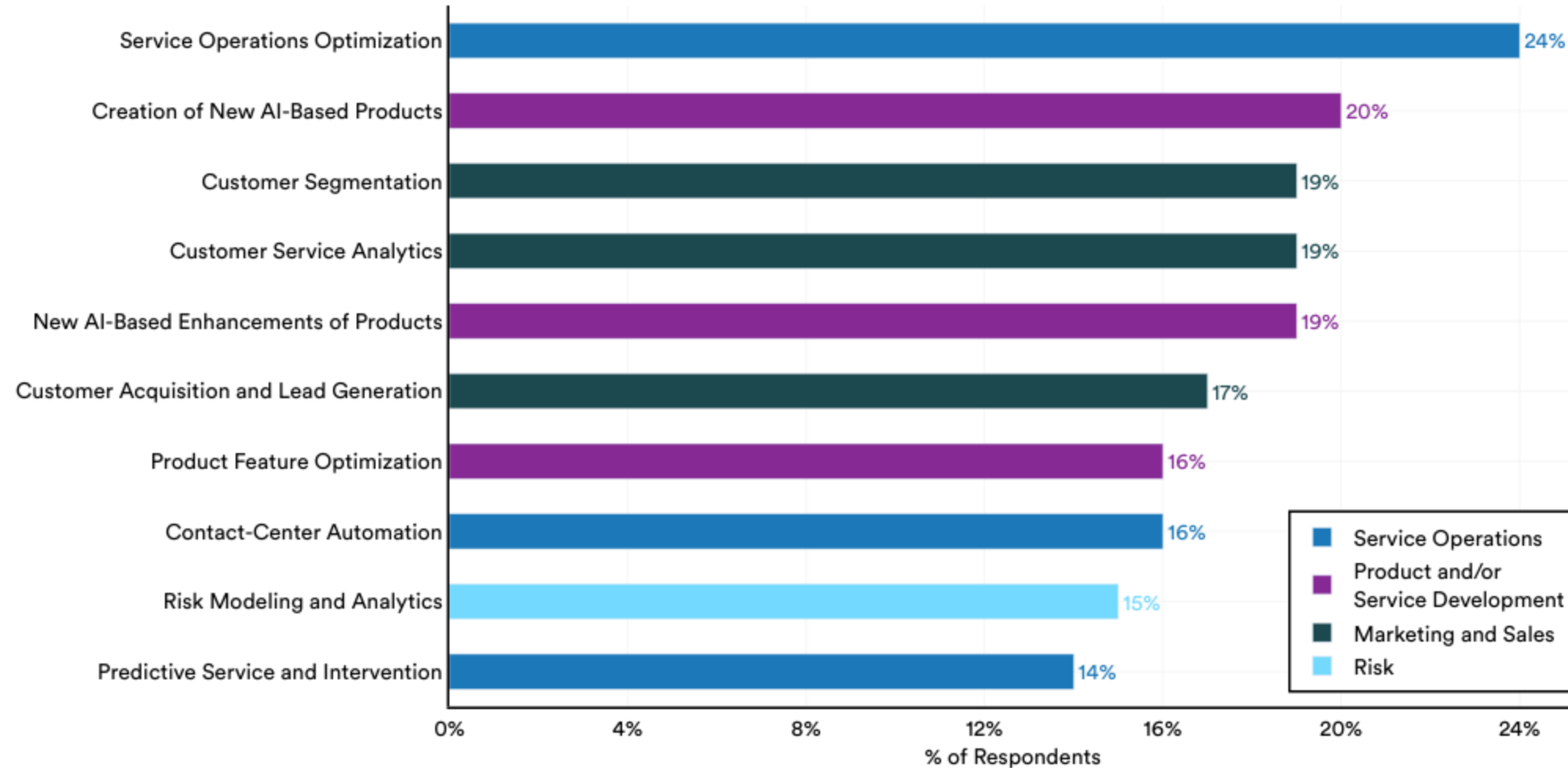
# Use-case by size

- Large enterprise focus on internal applications:
  - Cost-cutting, process automation, detecting fraud
- SME are customer-centric:
  - Satisfaction, experience, insights, customer churn

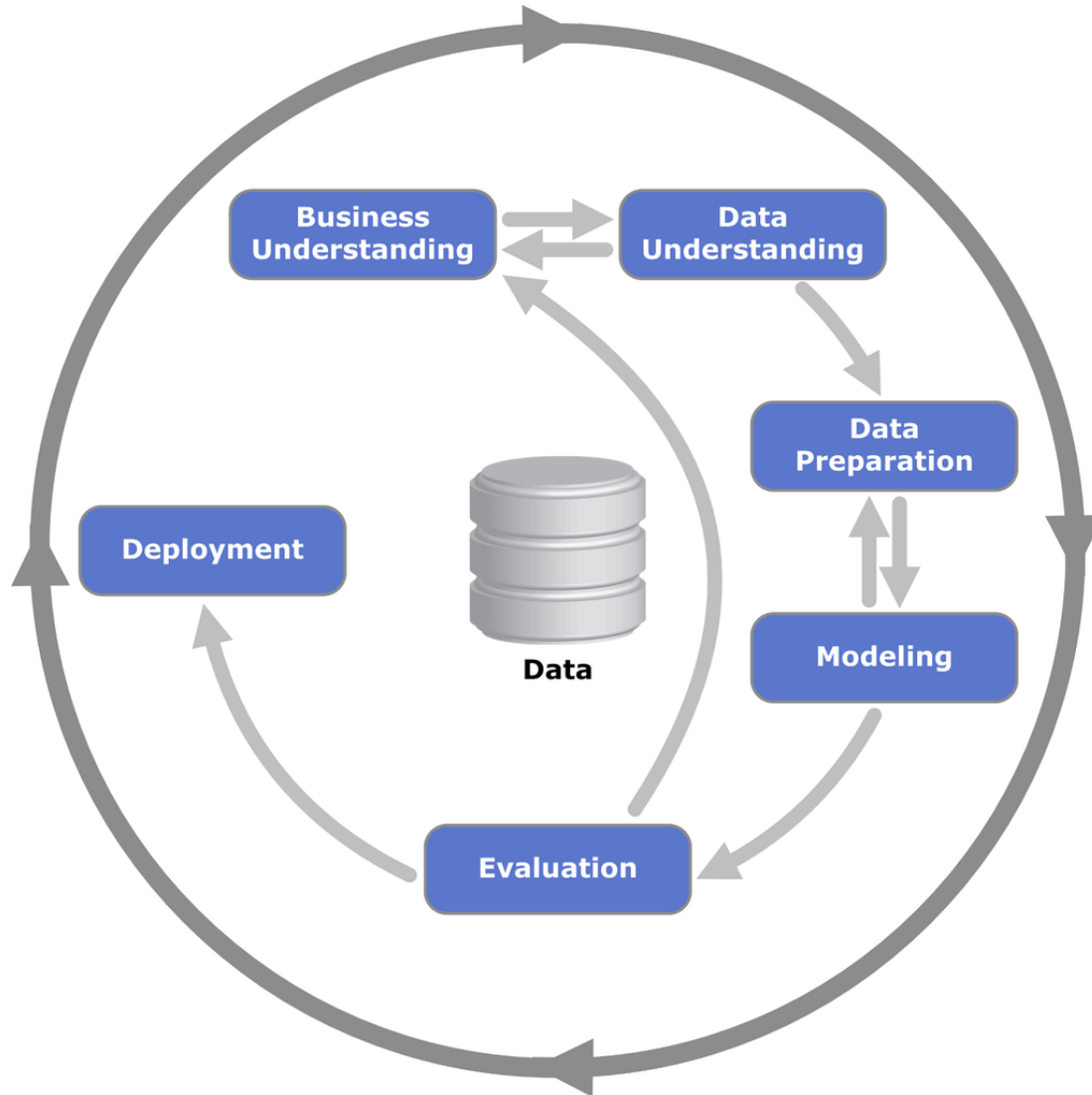
# AI use-cases

## Most Commonly Adopted AI Use Cases by Function, 2022

Source: McKinsey & Company Survey, 2022 | Chart: 2023 AI Index Report

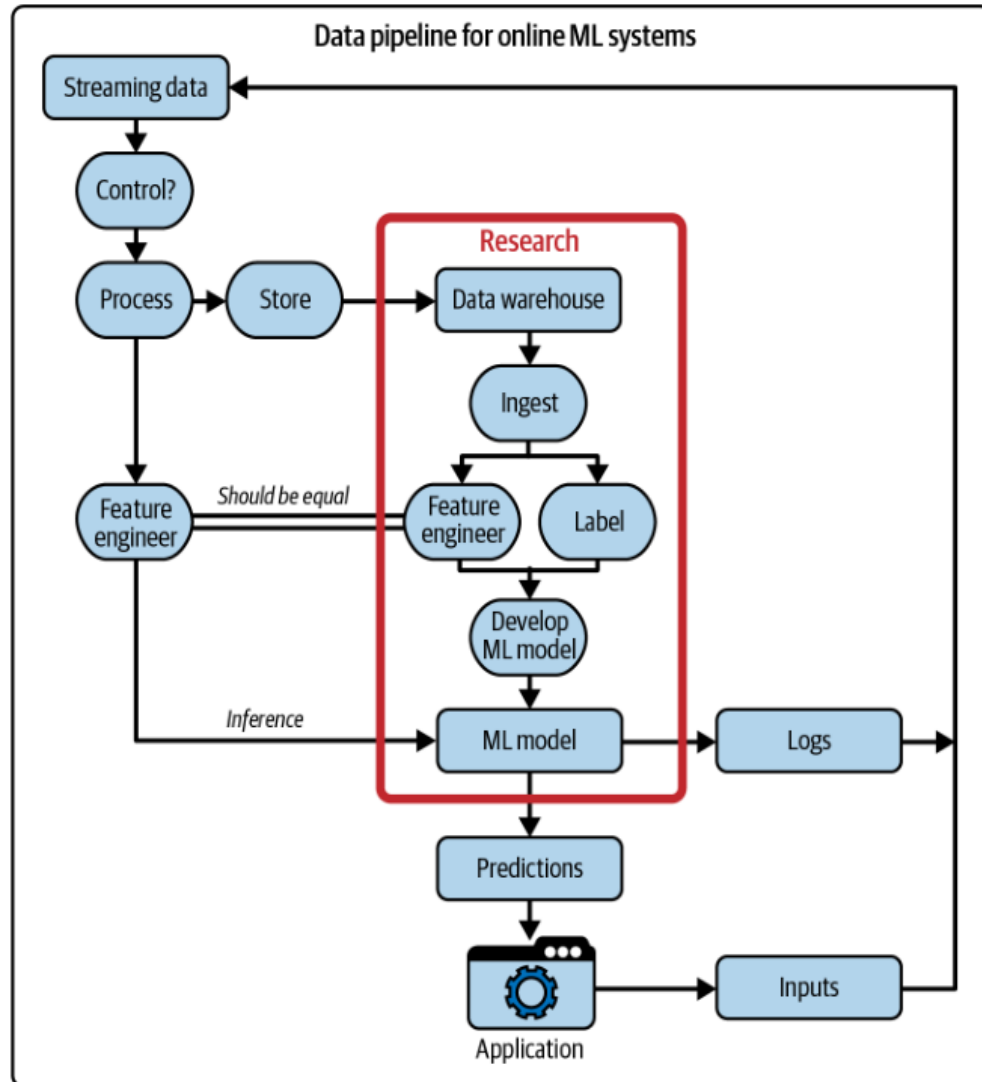


# Data mining



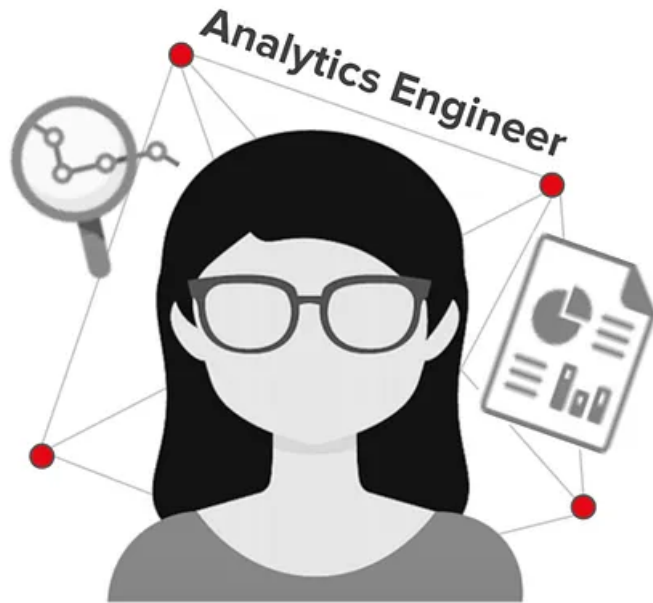
*CRISP-DM, Wikipedia*

# Online ML



*Designing Machine Learning Systems*

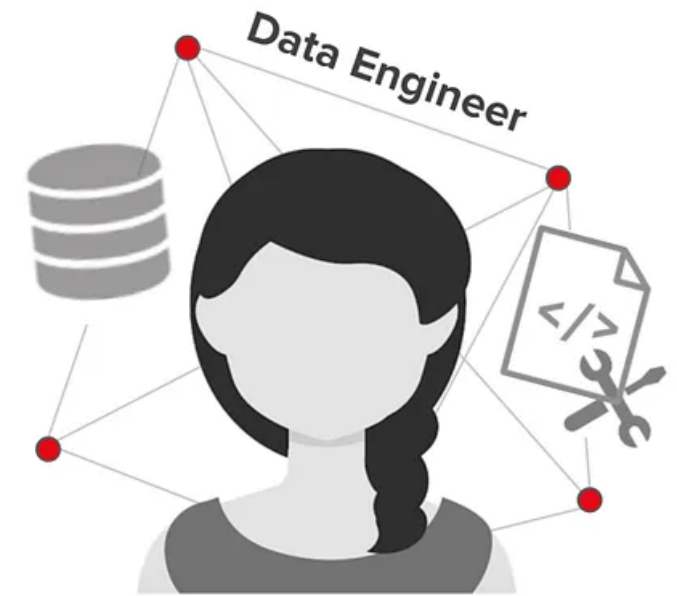
# Roles and tools



**tools:** Sublime, Atom, Tableau  
**languages:** SparkSQL, Presto, Python



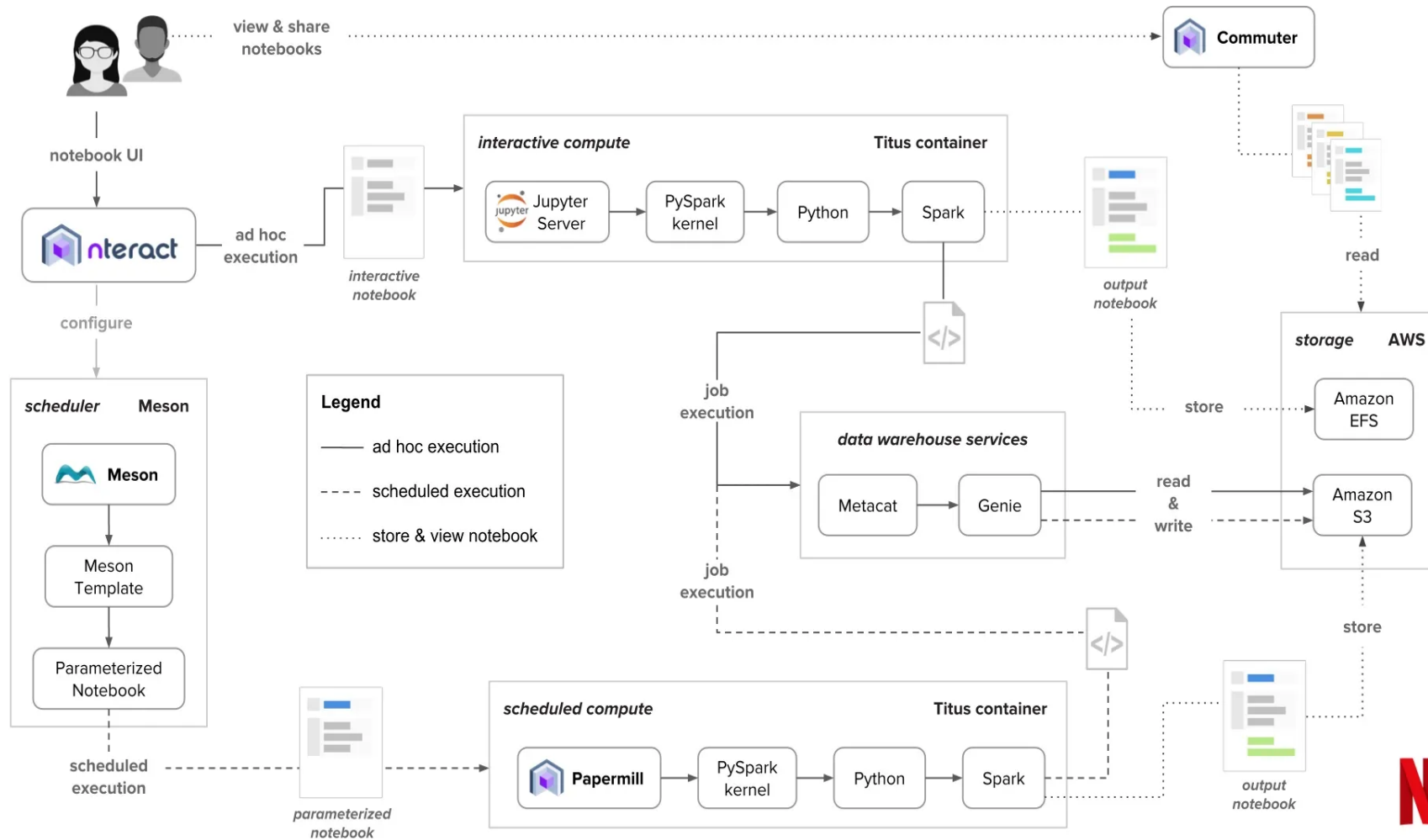
**tools:** Jupyter, RStudio, PyCharm  
**languages:** Python, Presto, R, PySpark



**tools:** IntelliJ, PyCharm, Sublime  
**languages:** Scala, Spark, Python, SQL

*Netflix Technology Blog*

# Infrastructure

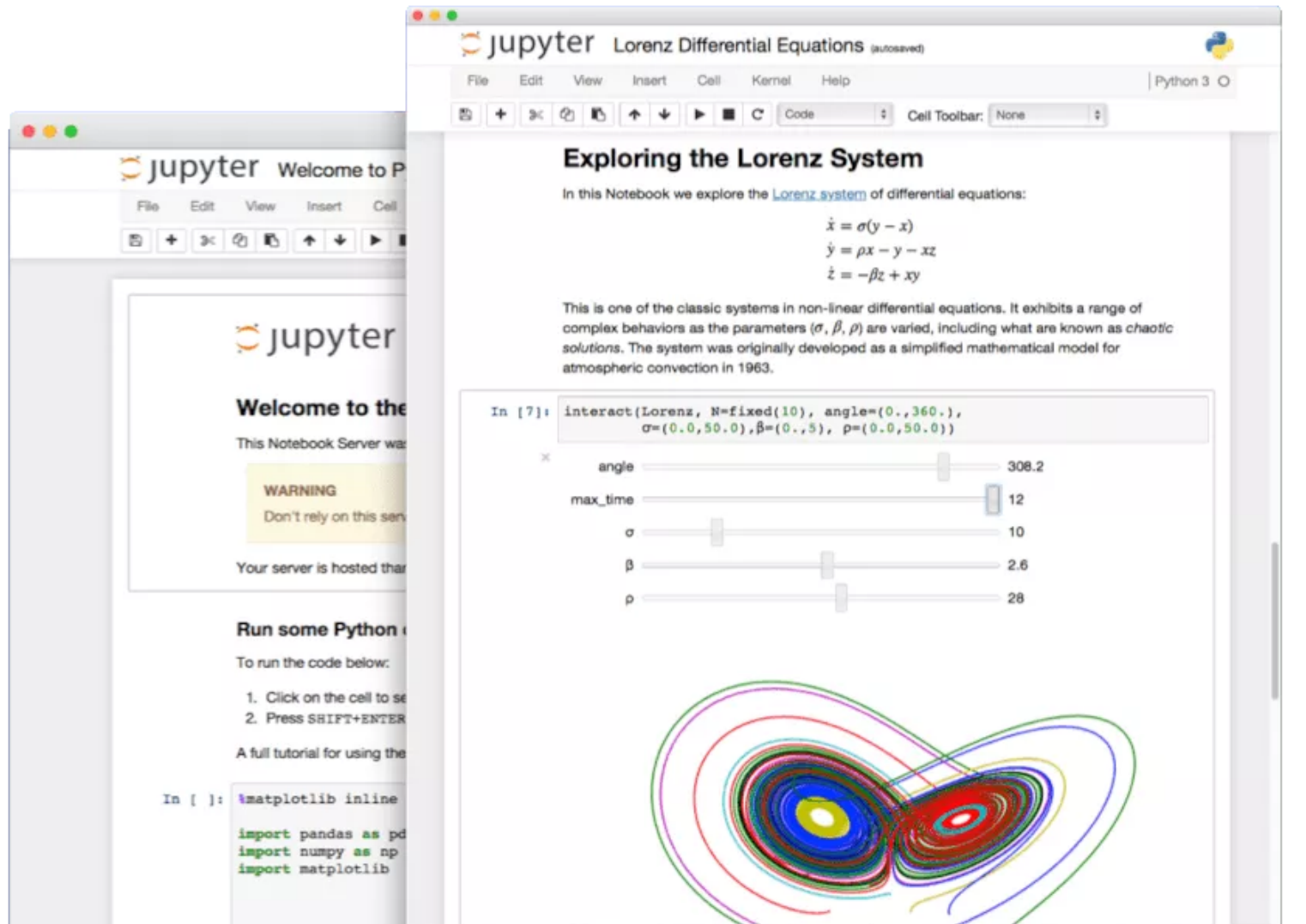


# Integration

- Database:
  - SQLAlchemy
- Cloud:
  - AWS SDK for Python
- Stream processing:
  - Faust vs. Apache Kafka / Spark / Storm / Samza / Flink
- GPU programming:
  - CUDA Python / PyCUDA / PyOpenCL



# Interactive prototyping



# Business intelligence and analytics


- Exploratory data analysis and visualisation
- Graphing libraries:
  - Matplotlib, seaborn
- Web app frameworks:
  - plotly, Dash, Flask
- Compatibility:
  - Tableau, Power BI

# Business intelligence



# Outsourcing and talent acquisition

## Results



**Passenger Screening Algorithm Challenge** ⋮


Improve the accuracy of the Department ...

Featured

518 Teams

---

**\$1,500,000** 6 years ago



**Zillow Prize: Zillow's Home Value Prediction...** ⋮


Can you improve the algorithm that chang...

Featured

3770 Teams

---

**\$1,200,000** 6 years ago



**Data Science Bowl 2017** ⋮


Can you improve lung cancer detection?

Featured

1972 Teams

---

**\$1,000,000** 7 years ago



**Vesuvius Challenge - Ink Detection** ⋮


Resurrect an ancient library from the ashe...

Featured · Code Competition

1249 Teams

---

**\$1,000,000** 8 months ago



**Heritage Health Prize** ⋮

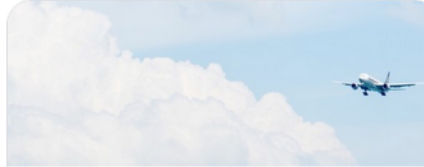
Identify patients who will be admitted to a...

Featured

1350 Teams

---

**\$500,000** 11 years ago



**GE Flight Quest** ⋮

Think you can change the future of flight?

Featured

172 Teams

---

**\$250,000** 11 years ago

# Real-world ML systems

<div> <div>Hide fields</div> <div>Filter</div> <div>Group</div> <div>Sort</div> <div></div> <div></div> </div>							
<input type="checkbox"/>	Company	Industry	Short Description (< 5 words)	Title	Tag	Year	Link
1	Stripe	Fintech and banking	Prevent fraudulent transactions	How we built it: Stripe Radar	fraud detection	2023	<a href="https://stripe.com/blog/how-we-built-it-stripe-radar">https://stripe.com/blog/how-we-built-it-stripe-radar</a>
2	Walmart	E-commerce and retail	Recommend complementary items	Personalized 'Complete the Look' model	recommender system product feature CV	2023	<a href="https://medium.com/walmartglobaltech/personalized-...">https://medium.com/walmartglobaltech/personalized-...</a>
3	Uber	Delivery and mobility	Forecast demand for airport rides	Demand and ETR Forecasting at Airports	demand forecasting	2023	<a href="https://www.uber.com/en-GB/blog/demand-and-etr-forecasting-at-...">https://www.uber.com/en-GB/blog/demand-and-etr-forecasting-at-...</a>
4	Pinterest	Social platforms	Prevent advertiser churn	An ML based approach to proactive advertiser churn ...	churn prediction	2023	<a href="https://medium.com/pinterest-engineering/an-ml-based-...">https://medium.com/pinterest-engineering/an-ml-based-...</a>
5	Stitch Fix	E-commerce and retail	Generate ad headlines	A New Era of Creativity: Expert-in-the-loop Generative AI at ...	product feature NLP generative AI	2023	<a href="https://multithreaded.stitchfix.com/blog/2023/03/06/expert-in-the-...">https://multithreaded.stitchfix.com/blog/2023/03/06/expert-in-the-...</a>
6	Swiggy	Delivery and mobility	Recommend items to order	Building a mind reader at Swiggy using Data Science	product feature recommender system	2023	<a href="https://bytes.swiggy.com/building-a-mind-reader-at-swiggy-using-...">https://bytes.swiggy.com/building-a-mind-reader-at-swiggy-using-...</a>
7	Microsoft	Tech	Diagnose production incidents with LLM	Large-language models for automatic cloud incident ...	ops generative AI product feature	2023	<a href="https://www.microsoft.com/en-us/research/blog/large-language-...">https://www.microsoft.com/en-us/research/blog/large-language-...</a>
8	Foodpanda	Delivery and mobility	Optimize menu sorting order	Menu Ranking	content personalization	2023	<a href="https://medium.com/foodpanda-data/menu-ranking-422ad21f381e">https://medium.com/foodpanda-data/menu-ranking-422ad21f381e</a>
9	Zillow	E-commerce and retail	Estimate the house market value	Building the Neural Zestimate	pricing	2023	<a href="https://www.zillow.com/tech/building-the-neural-zestimate/">https://www.zillow.com/tech/building-the-neural-zestimate/</a>
10	Airbnb	Travel E-commerce and retail	Identify user interests	Prioritizing Home Attributes Based on Guest Interest	NLP item classification search	2023	<a href="https://medium.com/airbnb-engineering/prioritizing-home-...">https://medium.com/airbnb-engineering/prioritizing-home-...</a>
11	GitHub	Tech	Generate code and code suggestions	Inside GitHub: Working with the LLMs behind GitHub Copilot	product feature generative AI	2023	<a href="https://github.blog/2023-05-17-inside-github-working-with-the-llm...">https://github.blog/2023-05-17-inside-github-working-with-the-llm...</a>

## LLMs

- **e-commerce chatbots** → [Improving Customer Support: LangChain based Helpdesk & LLM-Powered AI Chatbot](#)
- **Social media bots** → [AutoTweeter](#)
- **Supply chain analytics** → [Leveraging LLMs with LangChain for Supply Chain Analytics](#)

## Waiting time analysis

- **Delivery and ride-hailing platforms** → [DeeprETA: An ETA Post-processing System at Scale](#)

Uber uses ETAs to calculate fares, estimate pickup times, match riders to drivers, plan deliveries, and more

- **Wayfair** → [Delivery-Date Prediction](#)

what we had to do to cut the gap between promised and realized delivery times in half

## Predictive analytics

- **Zalando** → [Deep Learning Based Forecasting: A Case Study from the Online Fashion Industry](#)

Demand forecasting in the online fashion industry is particularly amenable to global, data-driven forecasting models

## Recommender systems

- **GrubHub** → [Online Learning for Recommendations at Grubhub](#)  
up to a 45x cost savings and a +20% metrics increase
- **Etsy** → [How We Built a Multi-Task Canonical Ranker for Recommendations at Etsy](#)  
we provide buyers recommendations to help them find that one special item that stands out to them

## Image recognition

- **Etsy** → [From Image Classification to Multitask Modeling: Building Etsy's Search by Image Feature](#)  
Searching by image is being rapidly adopted throughout the e-commerce industry



## Customer satisfaction

- **LinkedIn** → [Improving the customer's experience via ML-driven payment routing](#)

Our goal is to improve the subscription renewal member and customer experience by building and using an adaptive, ML-based, payment routing to replace the rule-based process.

- **Dropbox** → [Optimizing payments with machine learning](#)

At Dropbox, we found that applying machine learning to our handling of customer payments has made us better at keeping subscribers happily humming along.

## Fraud detection

- **BlaBlaCar** → [How we used machine learning to fight fraud at BlaBlaCar](#)

BlaBlaCar has implemented various fraud detection and fraud management systems, some of which include machine learning algorithms

**The End?**