Data Catalogs
Human Curation + Machine Learning

David Ebbels, Senior Sales Engineer

April 13, 2022

What is a Data Catalog?

- A repository of metadata on information sources across an organization
  - Search & discovery
  - Data governance & curation
  - Collaboration & analysis

- Catalogs a broad range of information assets
  - Data sets, tables, articles, reports, queries, visualizations, conversations

- Includes common functionality such as:

Business Glossary   Lineage   Catalog Pages   Search
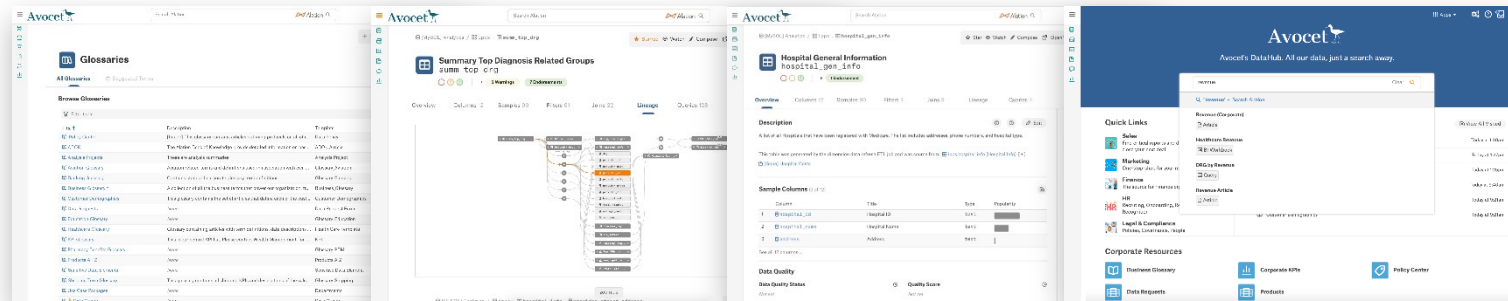
Answers these core questions:

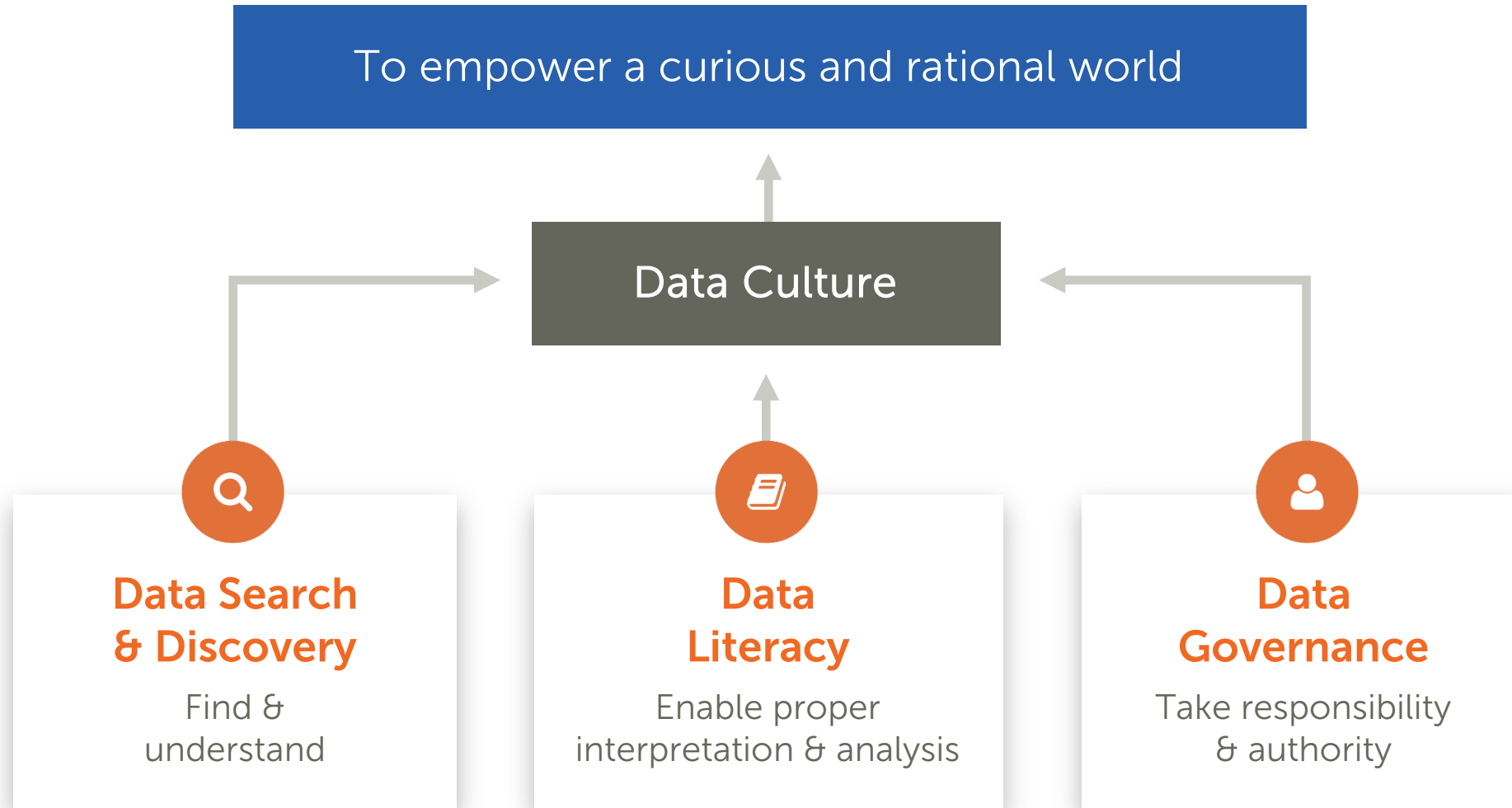**How to find information?**

Can it be used?

Should it be used?

How should it be used?

Alation

The Alation Vision

To empower a curious and rational world

Data Culture

Data Search & Discovery

Find & understand

Data Literacy

Enable proper interpretation & analysis

Data Governance

Take responsibility & authority

Alation

Human Curation + Machine Learning

Human Curation

Context & Understanding

Machine Learning

Scale & Documentation

Creating a Living Inventory

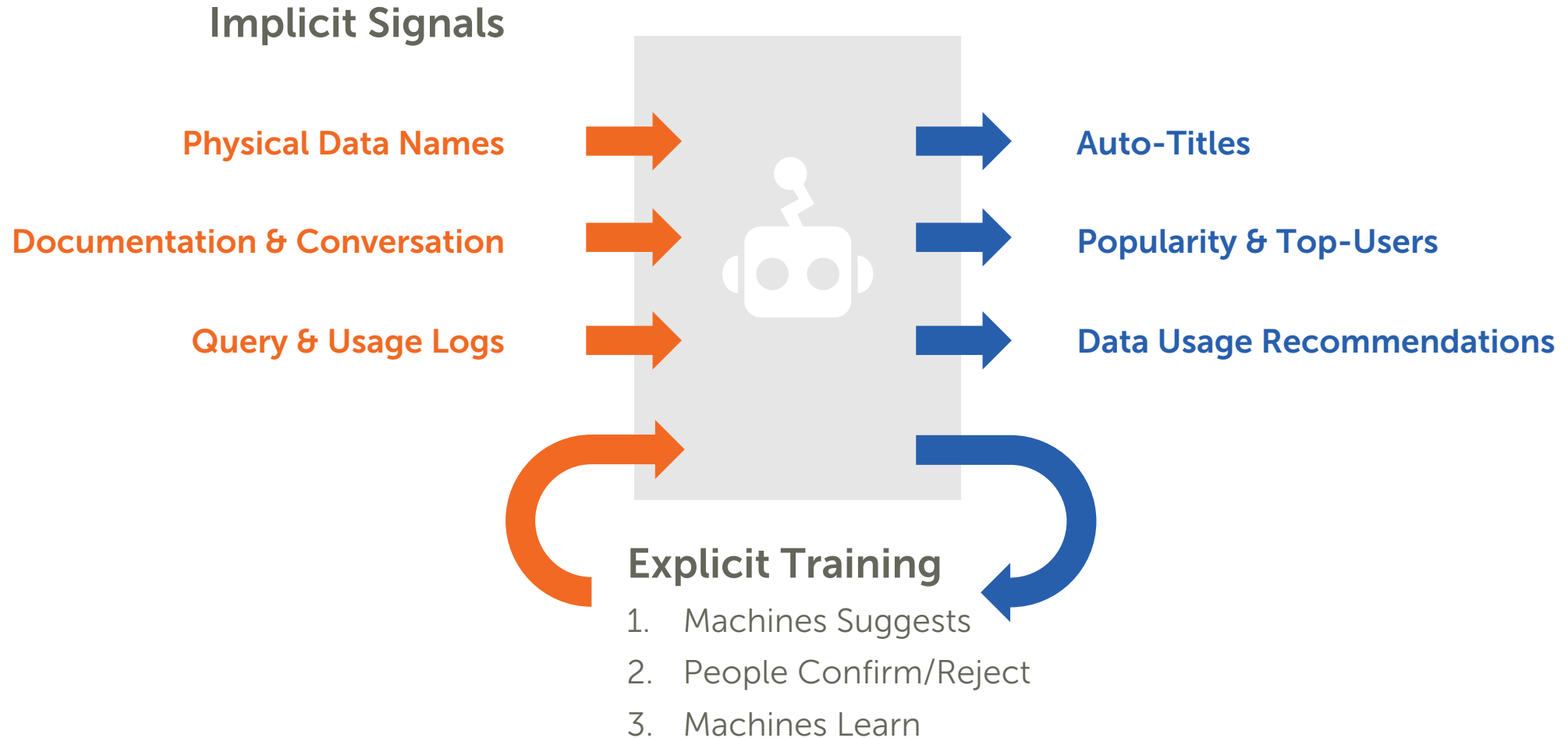Alation

Catalog Behavior I/O with Machine Learning

Driving better behavior in the future (output) by automatically learning from observed prior behavior (input). For example…

Implicit Signals

Physical Data Names → → **Auto-Titles**

Documentation & Conversation → → **Popularity & Top-Users**

Query & Usage Logs → → **Data Usage Recommendations**

Explicit Training

1. Machines Suggests
2. People Confirm/Reject
3. Machines Learn

Alation
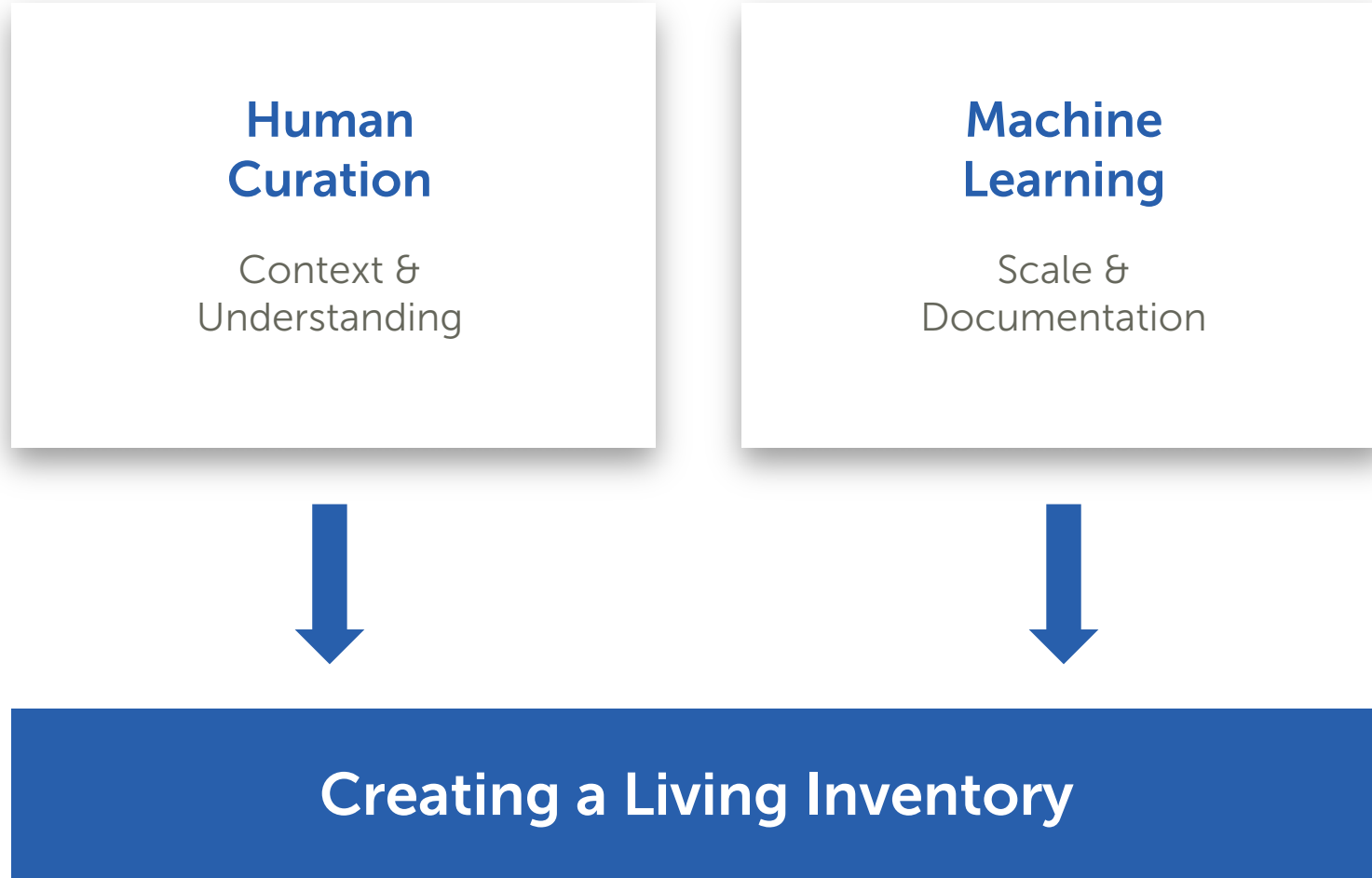
Demo

Learning from Users

Catalog Behavior I/O with Machine Learning

Driving better behavior in the future (output) by automatically learning from observed prior behavior (input). For example...
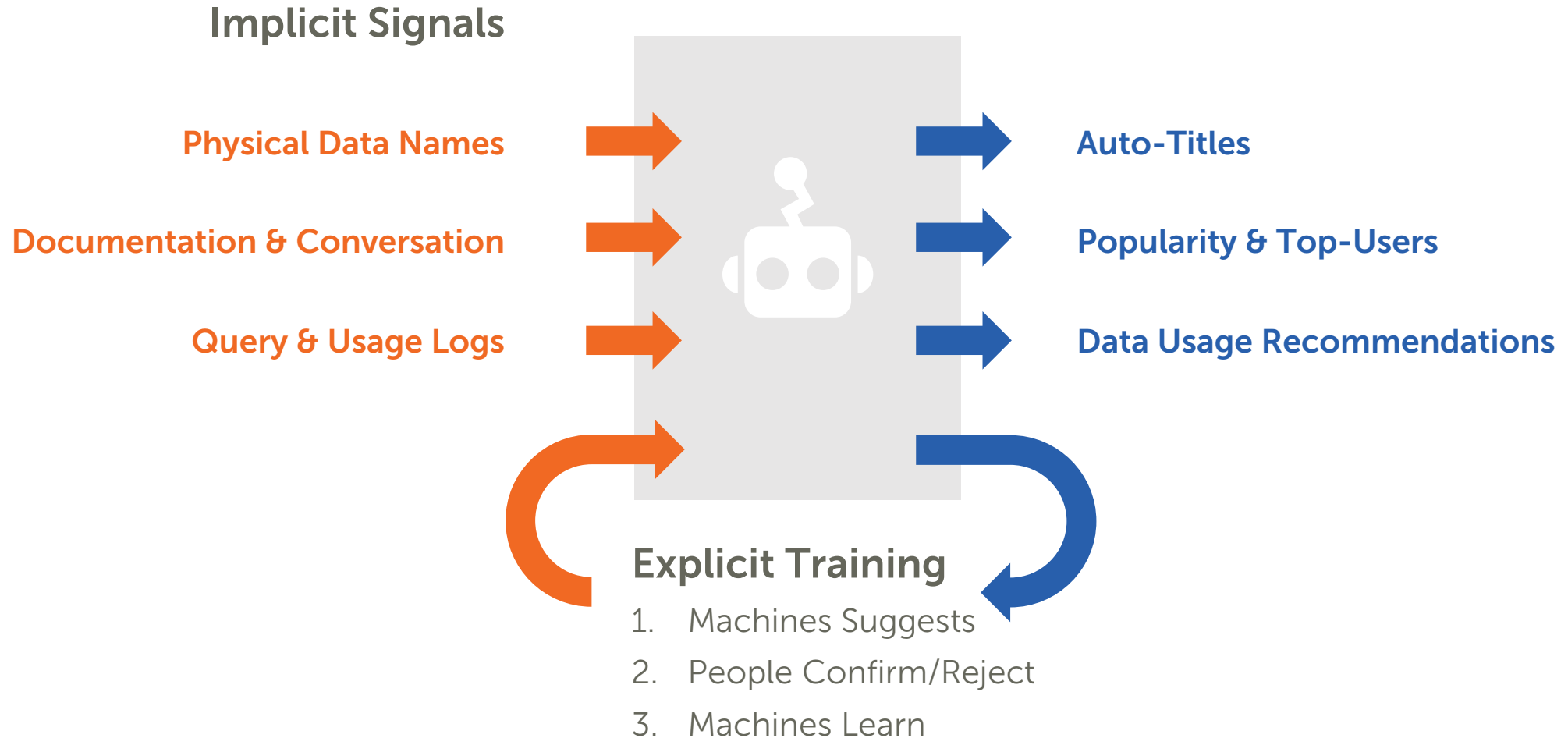
Implicit Signals

Physical Data Names → → **Auto-Titles**

Documentation & Conversation → → **Popularity & Top-Users**

Query & Usage Logs → → **Data Usage Recommendations**

Explicit Training

1. Machines Suggests
2. People Confirm/Reject
3. Machines Learn

Alation