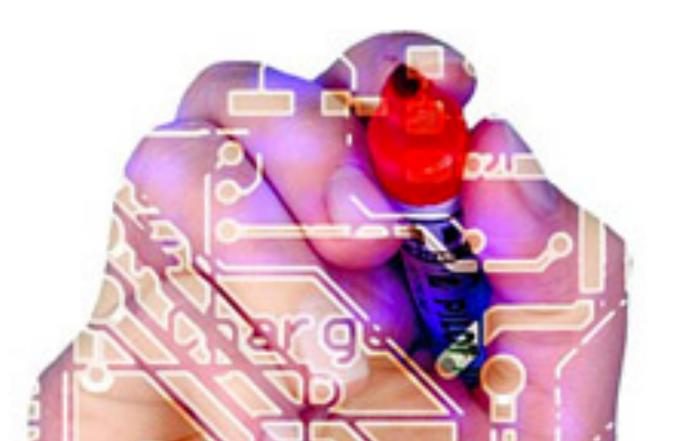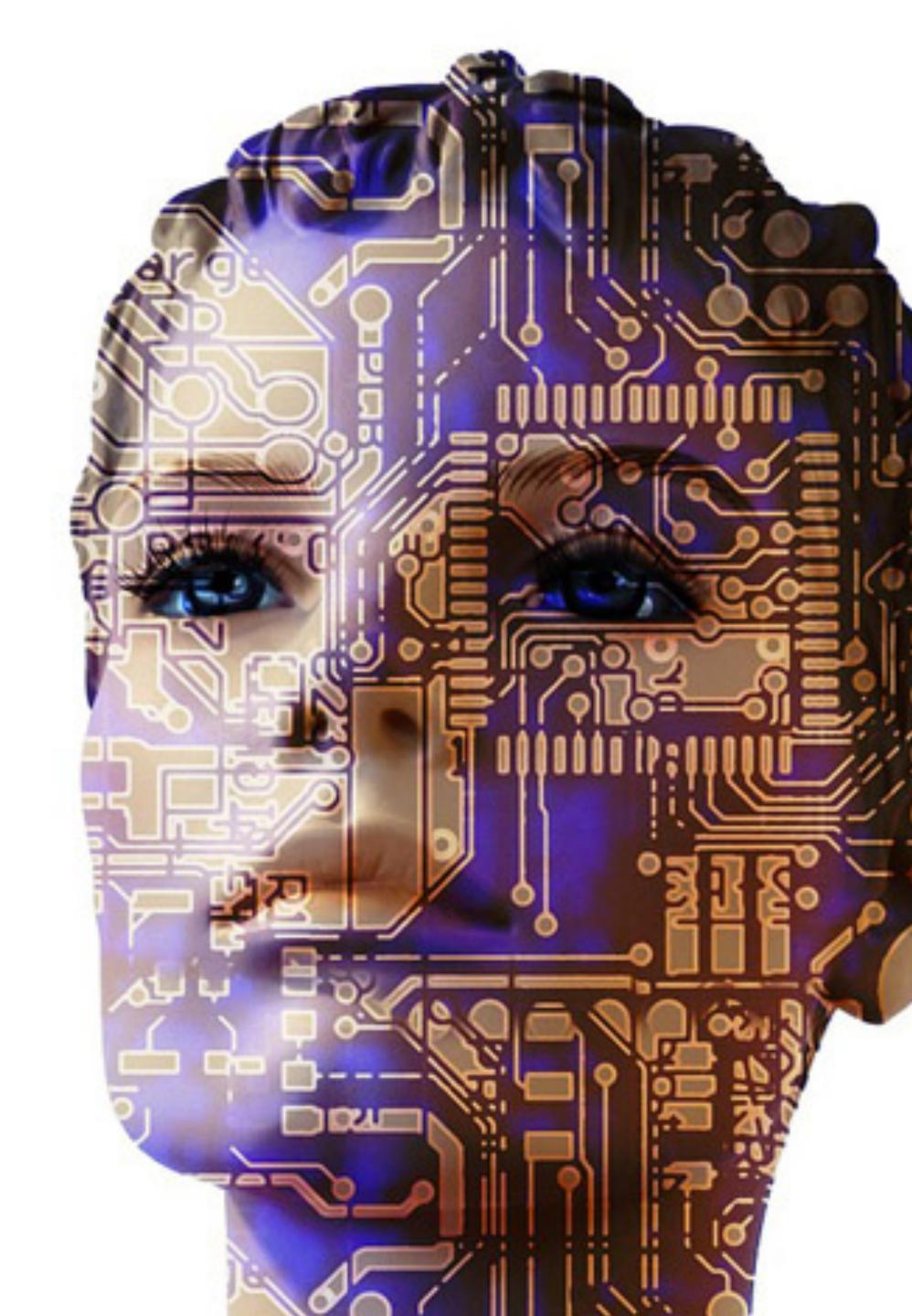# Hackathon 1: Q&A Engine Design

Fractal

What is the business context?

o Company X is a fortune-500 perfume manufacturing company

o It sells it products on Amazon

o It wishes to build an analytical query engine on top of the reviews and comments on amazon

o Sales and marketing team of company X would use this engine

o Based on brand/product perceptions on amazon Company X wishes to fine tune its

  o Sales and marketing strategy

  o New product innovation

Making AI-based Q&A engine for brand management and new product innovation

What is the technical problem statement?

o   Build a Q&A engine on top of amazon reviews/comments corpus; such that

o   If user searches with a product name/features; the engine

o   First finds out the most relevant review from all reviews in amazon

o   Given this first set of output, it then finds out from the rests the most diverse set of reviews

    o   Relevant

    o   Diverse

For example

o **Input question:** 'can I use 'B00028OSI0'on my face?'

o **Answers:**

  o 'All over! Buy it its worth every penny.' – positive sentiment

  o "Someone please tell me this product is alcohol free? It smells like alcohol and I'm not sure I should put it on my face. I do not see positive results yet either." – negative sentiment

  o "Initially I began but putting this on my face after washing it. I must say that it can't be used alone. ……." – negative sentiment

  o "The ingredients are non-toxic, non-drying and very helpful. It didn't work on my sensitive, picky, facial skin, but I do use it as a hair gel. …". – negative sentiment

Other details

- **Datasets:**

  - Amazon question/answer data on beauty products

  - Amazon product reviews on beauty products

- Validation Techniques:

  - Every reviews would be ranked based on

    - Relevance score – BM25 metrics

    - Sentiments

    - Both of these metrics must be shown on the final result set

Team structure and final deliverable

o **Team:**

   o Max 4 people per team

o Deliverables

   o Running web-based system (with/without UI part)

   o Code on public github (one account per team)

   o Technical presentations on approach (20 mins +5mins q&a)

   o Individual contributions must be explicitly mentioned in a slide

   o Free to refer anything available over web

Thank You!