

MBAi 448 | Winter 2026

Data Representations

Today

- Assignment 1 takeaways *10 minutes*
- Representations as tools *15 minutes*
- Representation case studies *20 minutes*
- Quick break *5 minutes*
- Ask an AI model *15 minutes*
- Assignment walkthrough *20 minutes*

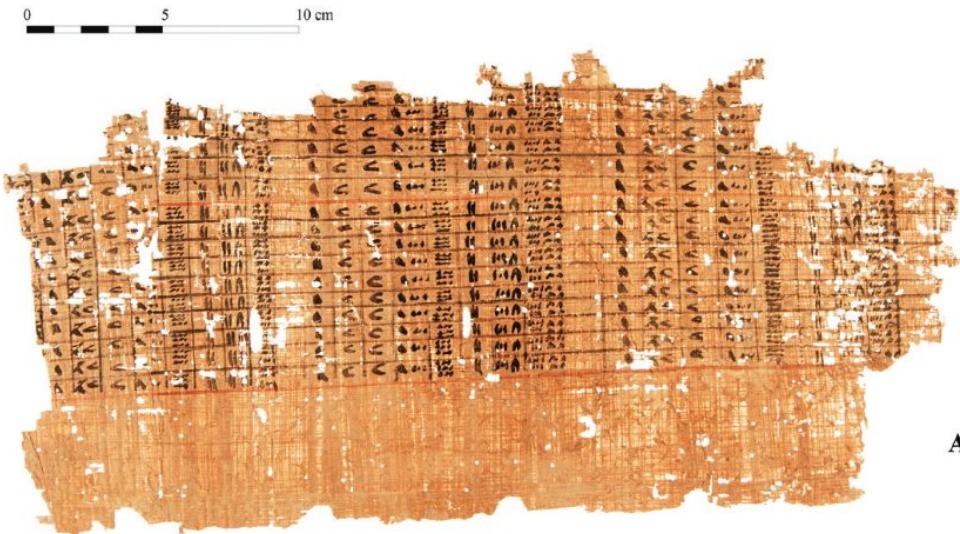
Representations as tools

Data Representations

Representations convey meaning to others



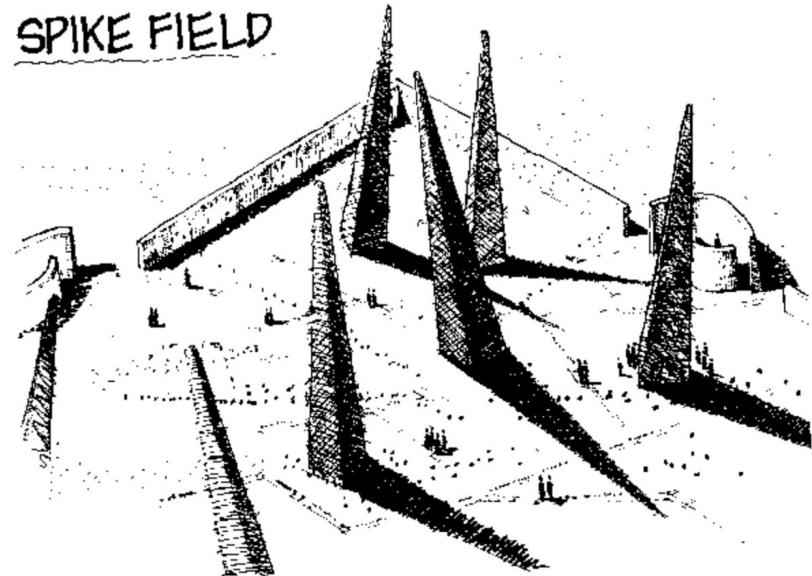
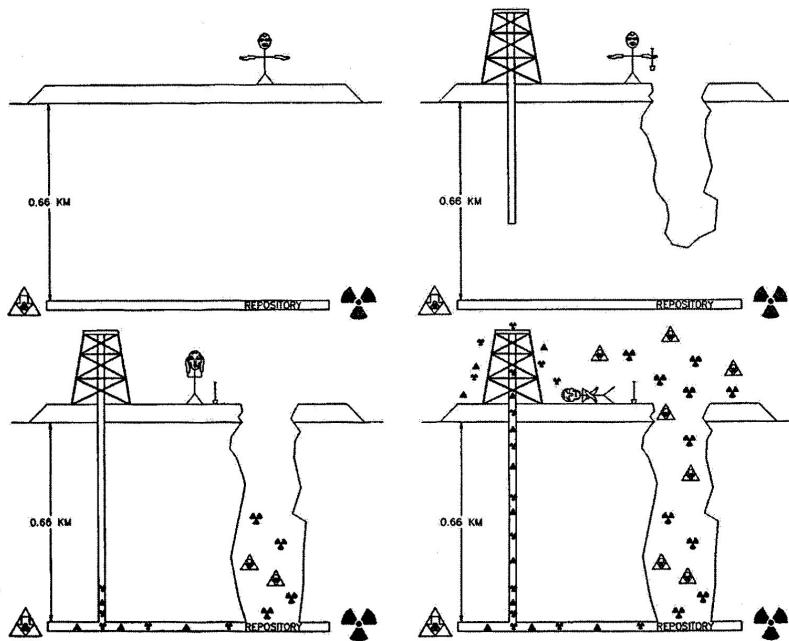
Representations facilitate work



Representations facilitate work



Representations are targeted at an audience



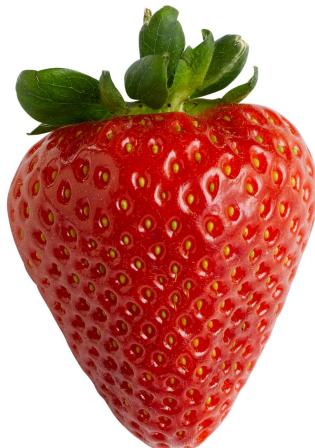
Representations are known through media



"And so the ['the medium is the message'] is intended to draw attention to the fact that a medium is not something neutral—it does something to people. It takes hold of them."

- Marshall McLuhan

Representations encode their meaning



“strawberry”

GPT-5.x & O1/3 GPT-4 & GPT-3.5 (legacy) GPT-3 (legacy)

strawberry

Clear Show example

Tokens	Characters
3	10

strawberry

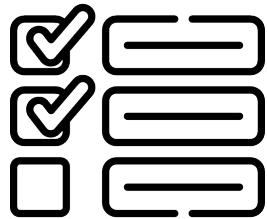
Representations can be self-referential



"The territory no longer precedes the map, nor survives it. Henceforth, it is the map that precedes the territory..."

- Jean Baudrillard

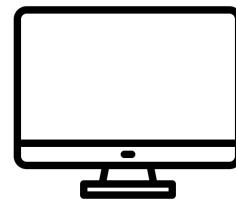
Representations are a tool



Work



Audience



Medium



Encoding

Representation case studies

Data Representations

Case Study: Google PageRank

Problem

- Early search engines relied on keywords and human-curated directories
- They were difficult to scale, easy to game, and often produced bad results

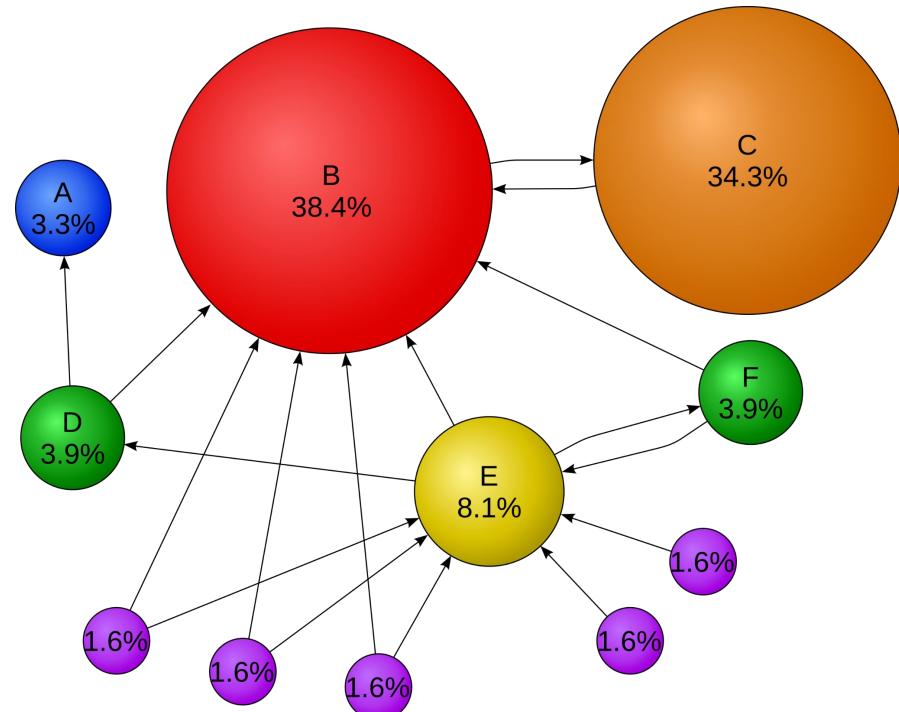
Representation

- PageRank

Case Study: Google PageRank

Representation

- PageRank
 - Work
 - Audience
 - Medium
 - Encoding



Case Study: Google PageRank

Enabling

- Web crawling and indexing
- Graph representations and algorithms
- Compute and storage infrastructure

Impact

-

Case Study: Spotify recommendations

Problem

- Radio channels lacked nuance and personal music curation was cumbersome
- Genre labels and other metadata could not deliver good recommendations

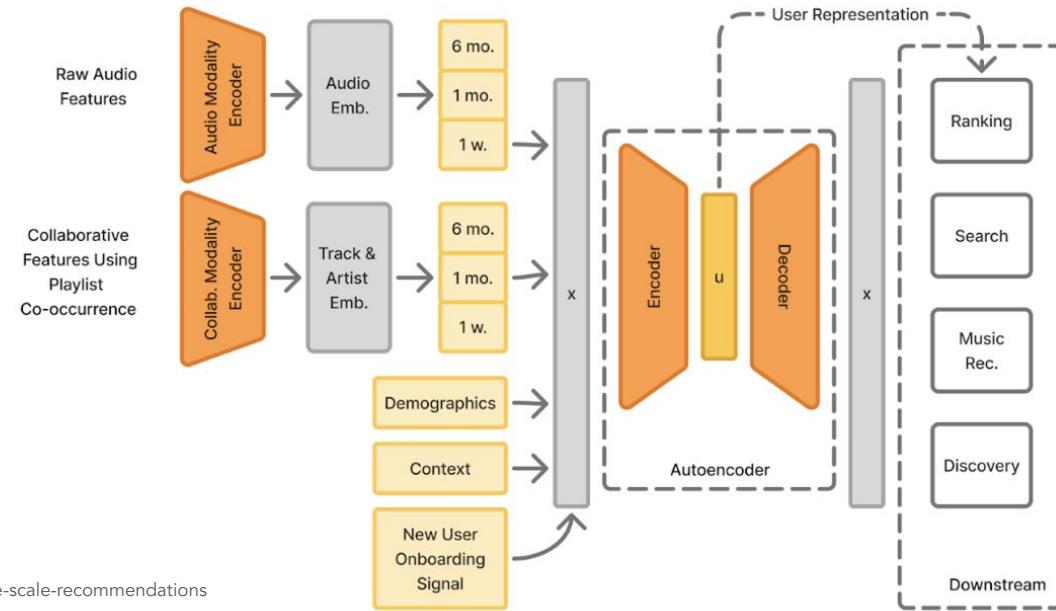
Representation

- High-dimensional embeddings for users, tracks, context, and behavior
 -

Case Study: Spotify recommendations

Representation

- High-dimensional embeddings for users, tracks, context, and behavior
 - Work
 - Audience
 - Medium
 - Encoding



Case Study: Spotify recommendations

Enabling

- Massive datasets of user behavior on the platform
- Recommendation algorithms and embedding representations
- Continuous user testing and data management

Impact

-

Case Study: Google Maps

Problem

- Spatial information is hard to communicate effectively through words
- Static maps do not capture important emergent transit developments

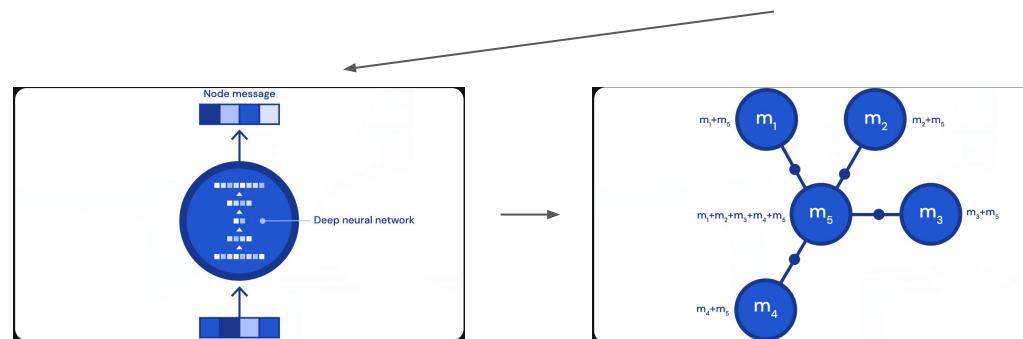
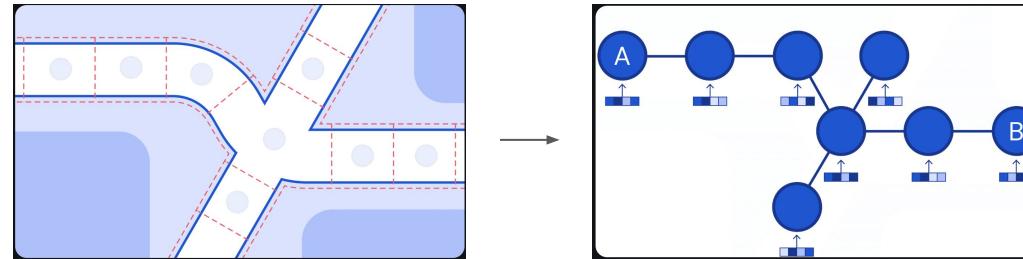
Representation

- Graph networks parameterized by user, location, environment, and time

Case Study: Google Maps

Representation

- Graph networks parameterized by user, location, environment, and time
 - Work
 - Audience
 - Medium
 - Encoding



Case Study: Google Maps

Enabling

- Massive amounts of historical, fine-grained movement data
- Real-time ingestion pipelines for aggregating and synthesizing new data
- Graph representations and algorithms

Impact

-

Ask an AI model

Data Representations

Getting AI to think with you, not for you



Stay mindful of bias



Question solutions



Expand your perspectives



**Spot weaknesses or
vulnerabilities**



**Start with your own
ideas**

Assignment walkthrough

Data Representations

