

# Data Science and Privacy in the EdTech Space

Hai-Ching Yang, Coursera Inc.  
Vinod Bakthavachalam, Coursera Inc.

**coursera**



**coursera**

Catalog

Search catalog



For Enterprise

Log In

Sign Up

Take the world's best courses,  
online.

Join for free



Stanford

UC San Diego

Duke

See All

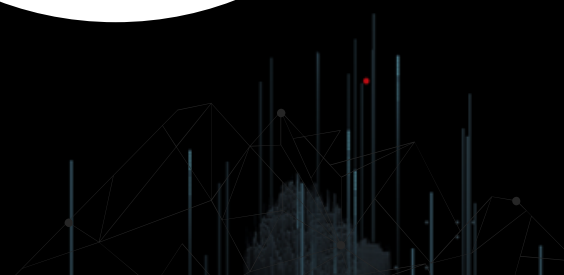
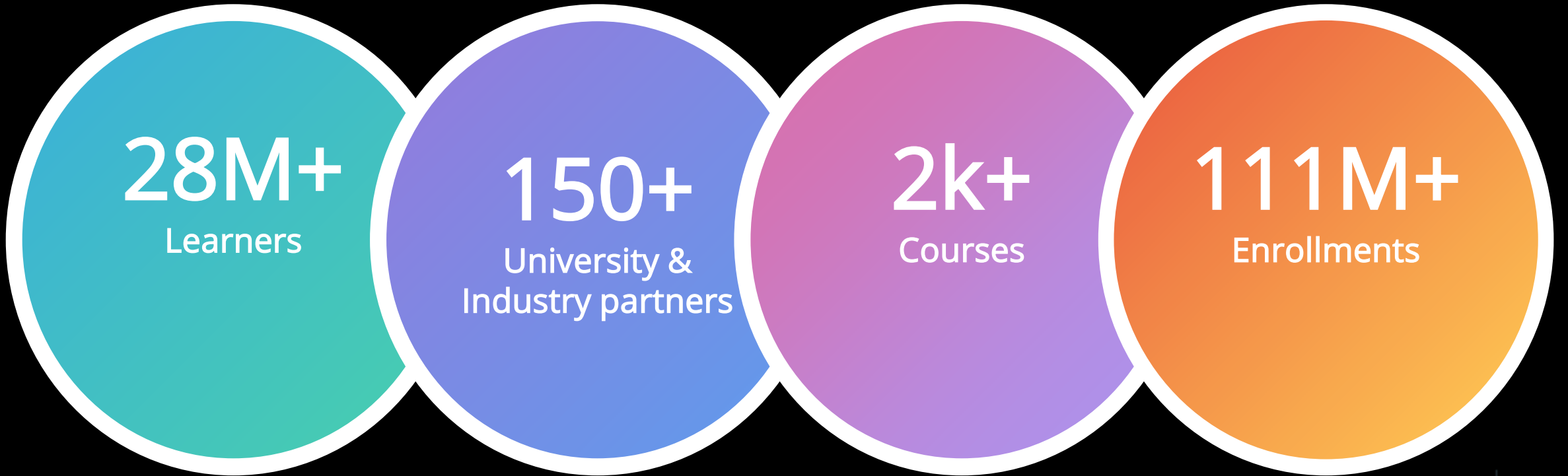
Master Cutting Edge AI

Andrew Ng's Deep Learning Specialization is now open!

Enroll Today



# Coursera is the largest online skill development platform



◀ With an active & global learner base





# Data Science at Coursera

We use data science to improve education

## DECISION SCIENCE

Draw out the user voice using statistical analysis to inform product direction.

## DATA PRODUCTS

Build data products to improve the learning experience and match users with the right solution for their goals.

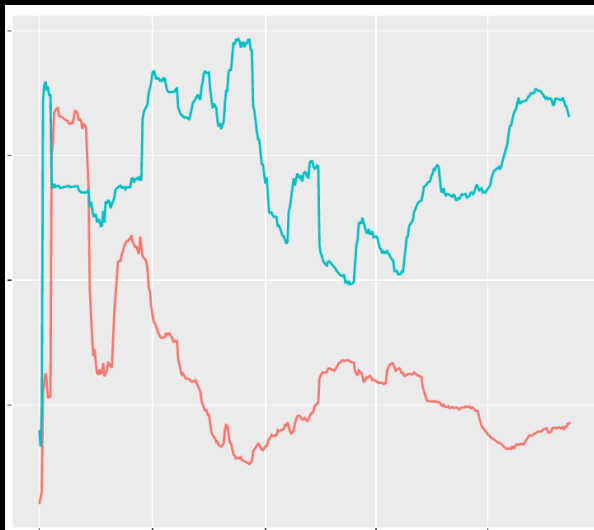
## INFRASTRUCTURE

Provide data resources and tools to enable and accelerate our product vision.

# Decision Science: draw out the user voice in the data

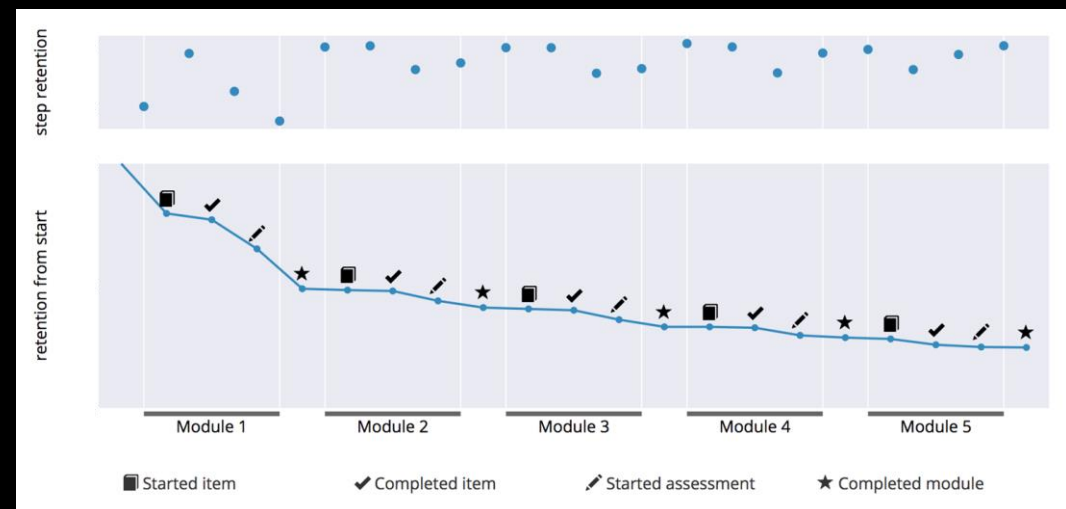
## Define Target Metrics

Standardize key metrics to track learner performance and understand levers that drive it



## Understand Learners

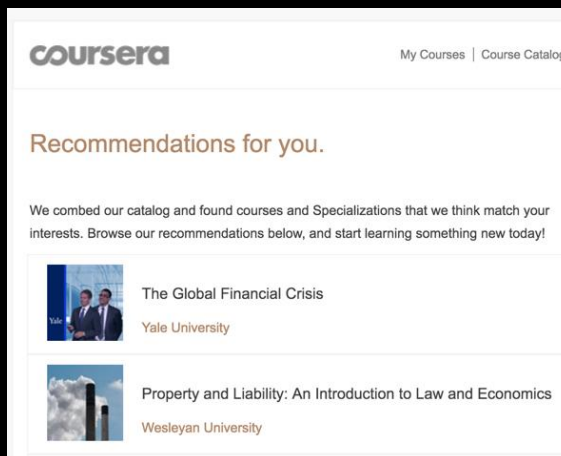
Perform analysis to highlight pain points in the learning process and develop solutions



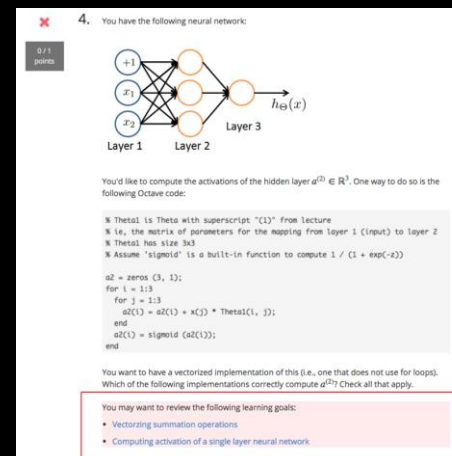
# Build data products that create value for learners

**Personalize learning**  
Recommend the optimal solution for learners based on their preferences, goals, background, etc.

**Drive mastery**  
Design the best learning experience that encourages learners to master difficult material



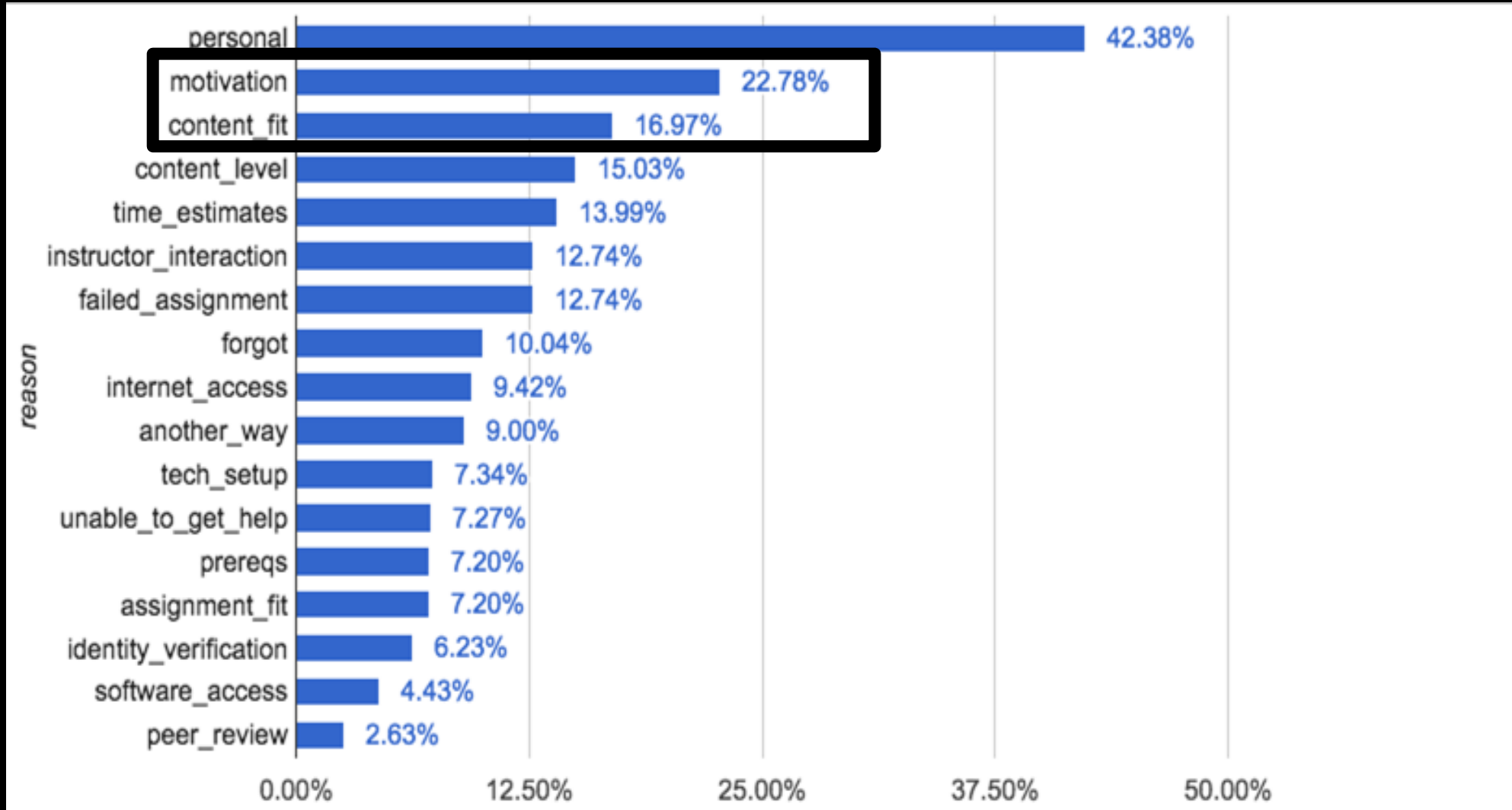
The screenshot shows the Coursera interface with a 'Recommendations for you' section. It lists two courses: 'The Global Financial Crisis' by Yale University and 'Property and Liability: An Introduction to Law and Economics' by Wesleyan University. The Coursera logo and navigation links are visible at the top.



The screenshot shows a problem statement for a neural network. It includes a diagram of a neural network with three layers: Layer 1 (input nodes  $x_1, x_2$  and bias  $+1$ ), Layer 2 (hidden nodes), and Layer 3 (output node  $h_{\Theta}(x)$ ). Below the diagram is Octave code for computing the activations of the hidden layer  $a^{(2)} \in \mathbb{R}^3$ . The code includes comments and a loop for computing the activations. A red box highlights the learning goals: 'Vectorizing summation operations' and 'Computing activation of a single layer neural network'.

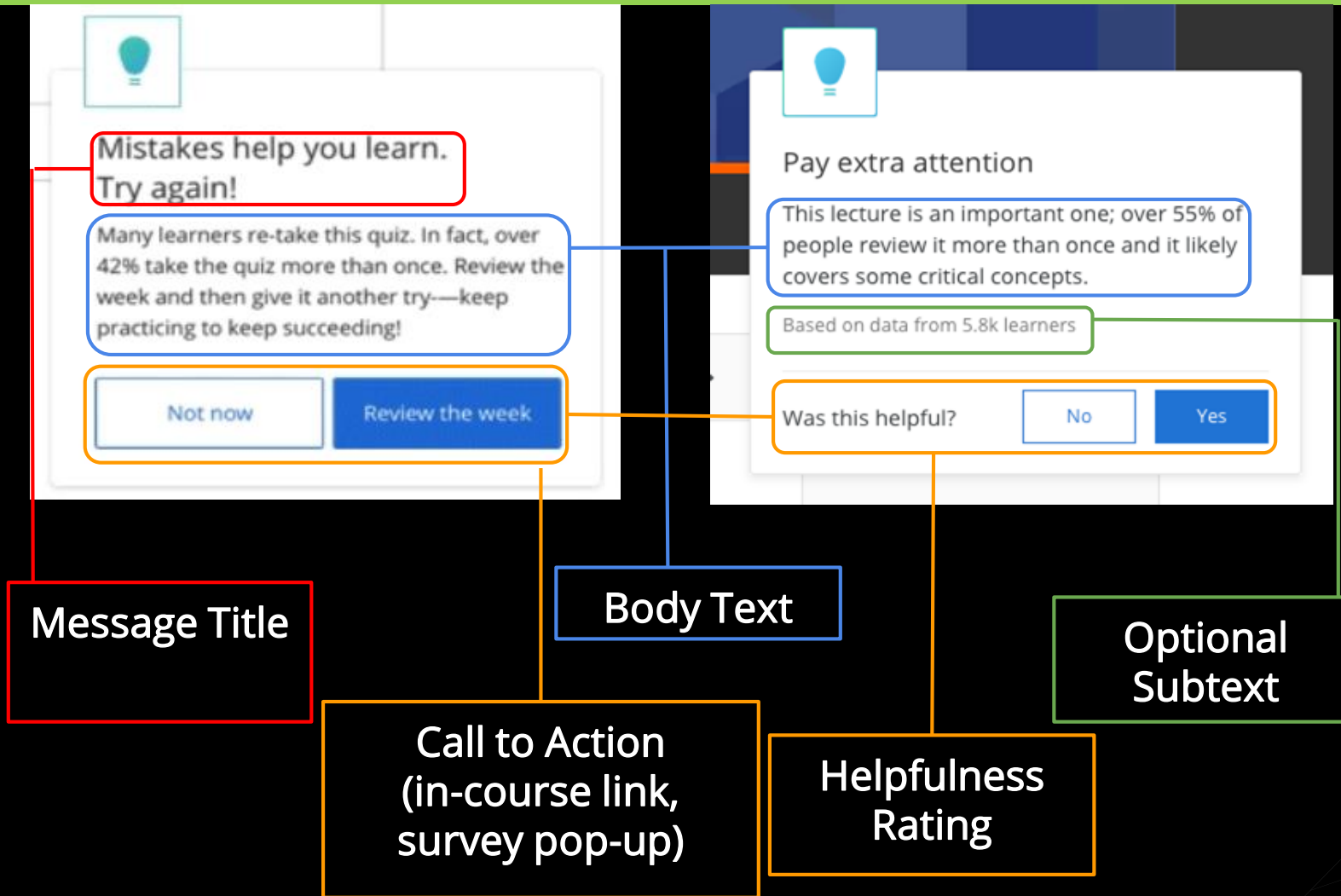


# Top Reasons for Dropping out of Courses



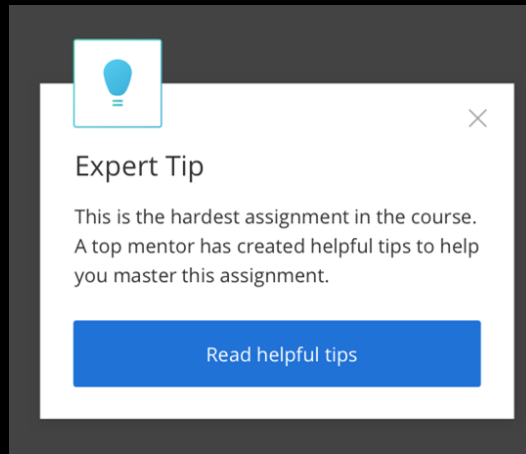


# How can we motivate learners → Situational Messaging



# How can we motivate learners → Situational Messaging

Imagine guidance options for  
a single difficult assignment



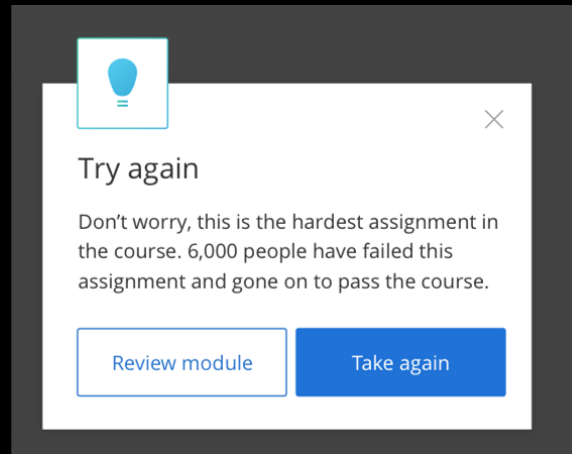
Expert Tip

This is the hardest assignment in the course. A top mentor has created helpful tips to help you master this assignment.

Read helpful tips

This notification features a lightbulb icon in a blue square at the top left and a close button (X) at the top right. The text is centered, and the button is a solid blue rectangle at the bottom.

*Tips before a learner starts*



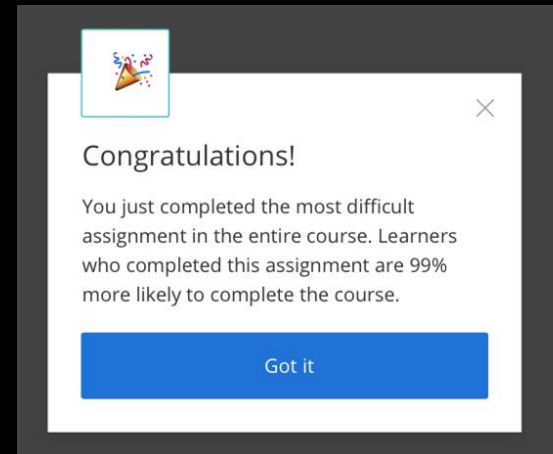
Try again

Don't worry, this is the hardest assignment in the course. 6,000 people have failed this assignment and gone on to pass the course.

Review module Take again

This notification features a lightbulb icon in a blue square at the top left and a close button (X) at the top right. The text is centered. At the bottom, there are two buttons: a white button with a blue border labeled 'Review module' and a solid blue button labeled 'Take again'.

*Support if a learner stumbles*



Congratulations!

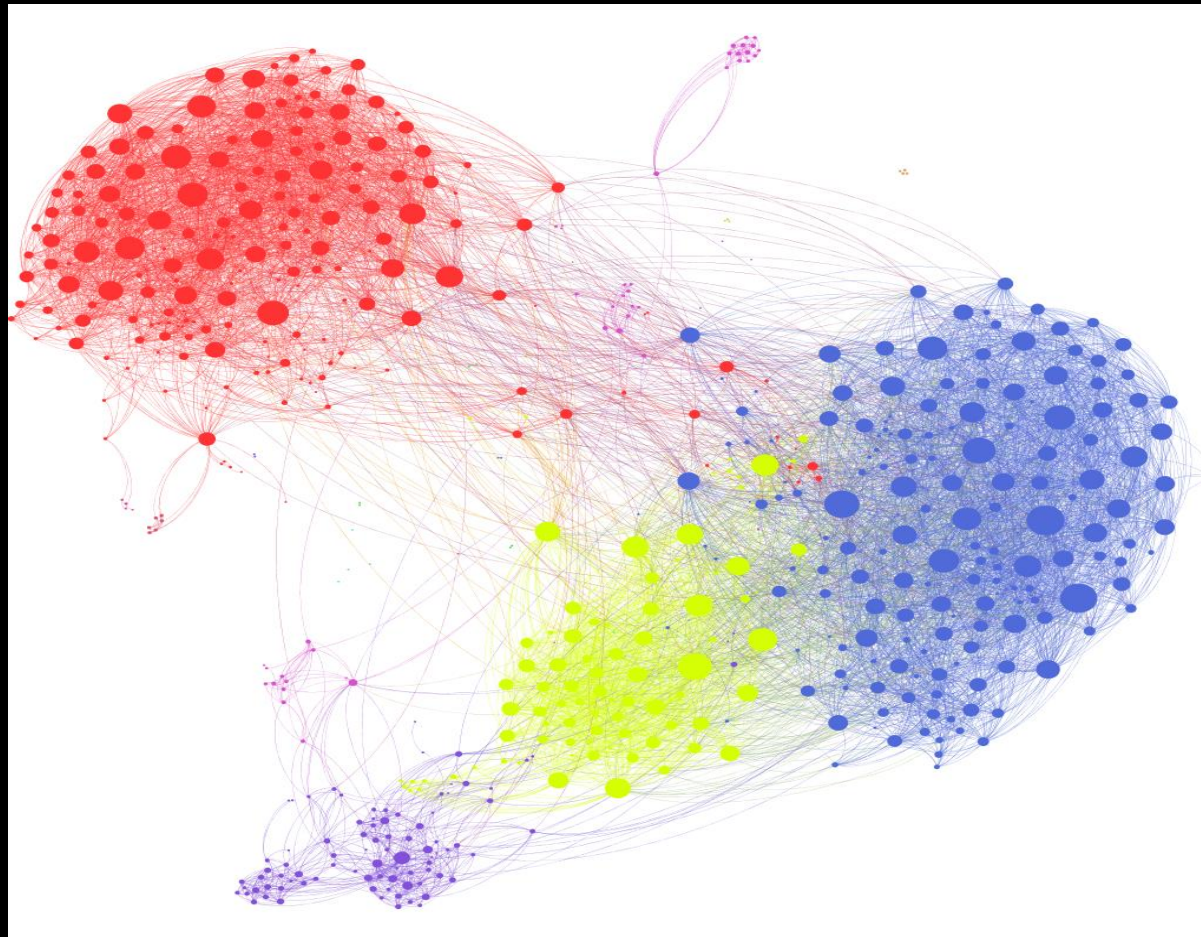
You just completed the most difficult assignment in the entire course. Learners who completed this assignment are 99% more likely to complete the course.

Got it

This notification features a party popper icon in a blue square at the top left and a close button (X) at the top right. The text is centered, and the button is a solid blue rectangle at the bottom.

*Motivation if a learner succeeds*

# How can we improve content fit? → Personalize Learning

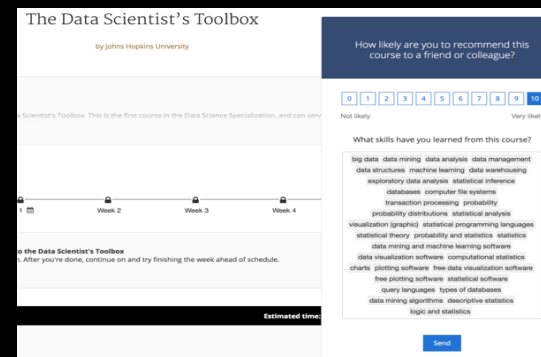


# Develop a common currency across content in skills

- Initial skill taxonomy and hierarchy from Wikipedia



- Tag skills to courses using machine learning and user voice



- Tagging model
- # of lexical occurrences in
- course title
  - course video transcripts
  - module descriptions
  - assessments

# Use this common currency to make personalized recommendations

Great! What are some skills that would help you advance your career?

You can choose topics in up to 3 of these categories:

Arts and Humanities   Business   Computer Science   **Data Science**

<input checked="" type="checkbox"/> Machine Learning	<input type="checkbox"/> Data Analysis	<input type="checkbox"/> Probability and Statistics
<input type="checkbox"/> Big Data	<input type="checkbox"/> Python Programming	<input checked="" type="checkbox"/> R Programming
<input type="checkbox"/> Business Analytics	<input checked="" type="checkbox"/> Data Visualization	<input type="checkbox"/> SQL
<input type="checkbox"/> Database	<input type="checkbox"/> Data Mining	<input type="checkbox"/> Recommender Systems

Life Sciences   Math and Logic   Personal Development

Physical Science and Engineering   Social Sciences   Language Learning



< My Coursera

Goal: Learn Machine Learning ...

**How you'll reach your goal with Coursera:**

- Explore your course recommendations (endorsed by industry experts).
- Customize your program by hiding courses that cover familiar material.
- Start learning! Work as fast or as slow as you like—you're in control, and we're here to support you.

**START LEARNING**

**A Crash Course in Data Science**  
Part of Executive Data Science Specialization. By now you have definitely heard about data science and big data. In this one-week class, we will provide a crash course in... [See Course Info](#) Enroll

**MACHINE LEARNING FOUNDATIONS • 3 COURSES**

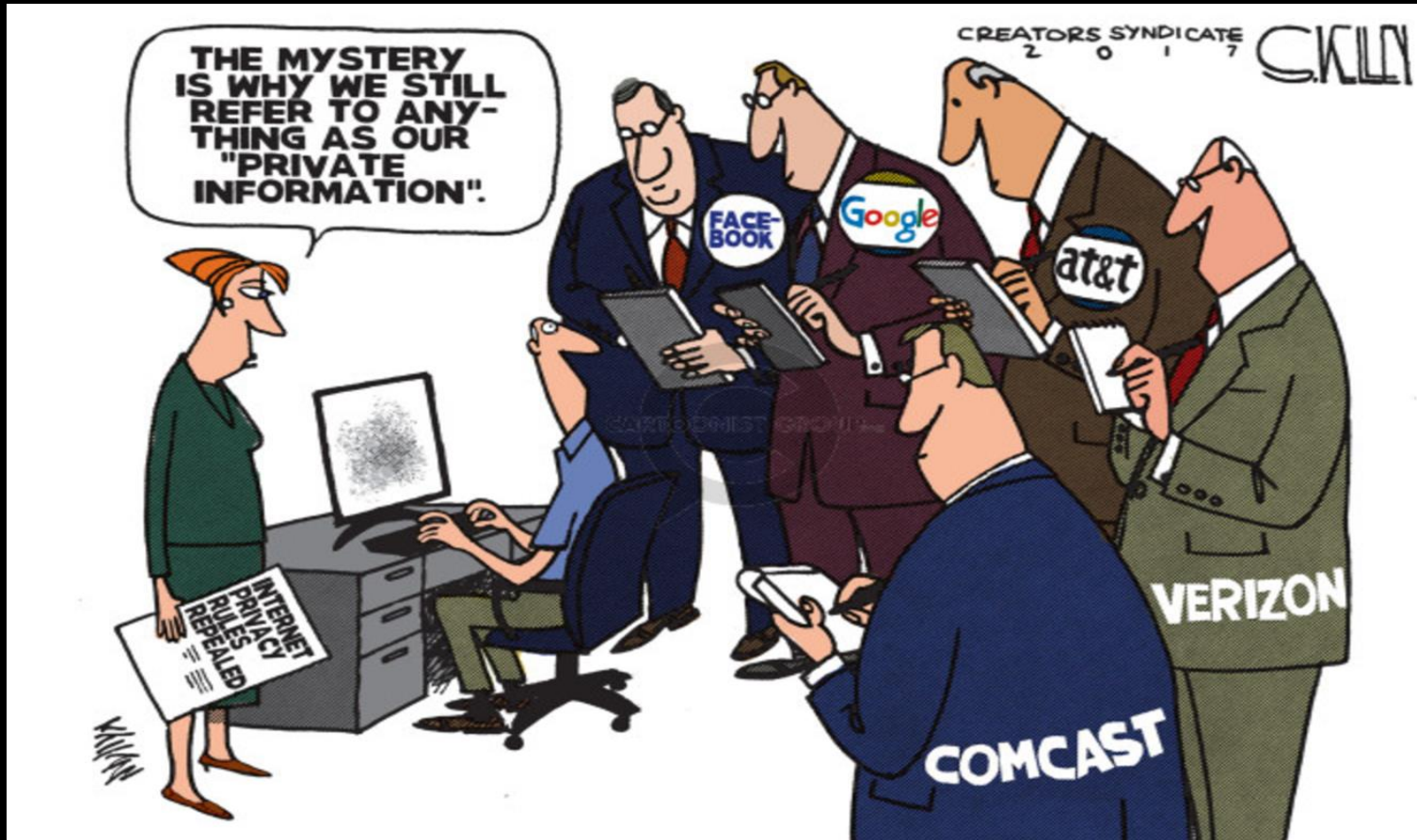
A Crash Course in Data Science (4 hours total) ...

Part of Executive Data Science Specialization • Johns Hopkins University

Introduction to Big Data (14 hours total) ...



# Does Online Privacy Still Exist?



Steve Kelley's Editorial Cartoons: <http://www.cartoonistgroup.com/store/add.php?id=156463>



## Why is Privacy Important for EdTech?



- As a rapidly scaling startup, Coursera deals with privacy concerns from learners, enterprise customers, university and industry partners, as well as government and nonprofits.
- Coursera has **28+ million** learners from around the world and we collect data to provide and improve the services and to conduct research.

**Key Q: How can we best ensure that our learners data is protected?**



## Case Hypo: Meet Anna, a Learner from France

Anna (age 15) from France is planning to take a data science specialization from University of Illinois on the Coursera platform.

What are the privacy concerns if she:

- signs up for an account?
- requests the information that Coursera collects?
- encounters a data breach situation?



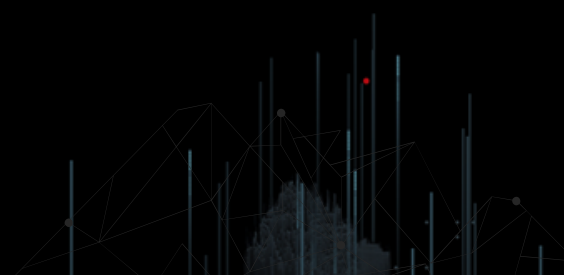




## What are we collecting from Anna?

- Non-Personally Identifiable Information
  - general info on which pages are visited, links clicked, when visitors come to our site.
- Personally Identifiable Information (PII)
  - any data that could potentially identify a specific individual.
  - e.g. name, email address

\*\*Unlike financial or health data, the information that EdTech companies collect are usually less sensitive. But even so, we must have adequate protections.





# What are international privacy concerns in dealing with a European learner like Anna?

## No Unified Privacy Regime

- Different countries have varying treatment of PII
- Hard for EdTech startups with global learners

## EU = Strict Regulation

- General Data Protection Regulation (GDPR) to be in effect in May 2018
- Affects all companies touching European personal data, heavy fines for violations

## Data Localization

- Some countries require data storage in the local country
- Burdensome and problematic for startups

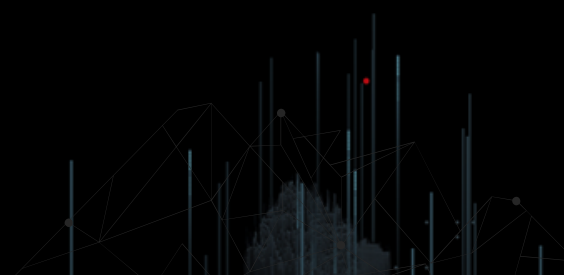


## Major US Privacy Regulations for EdTech Companies

US government heavily regulates the education industry, including:

- Children Online Privacy Protection Act (COPPA)
- Family Educational Rights and Protection Act (FERPA)

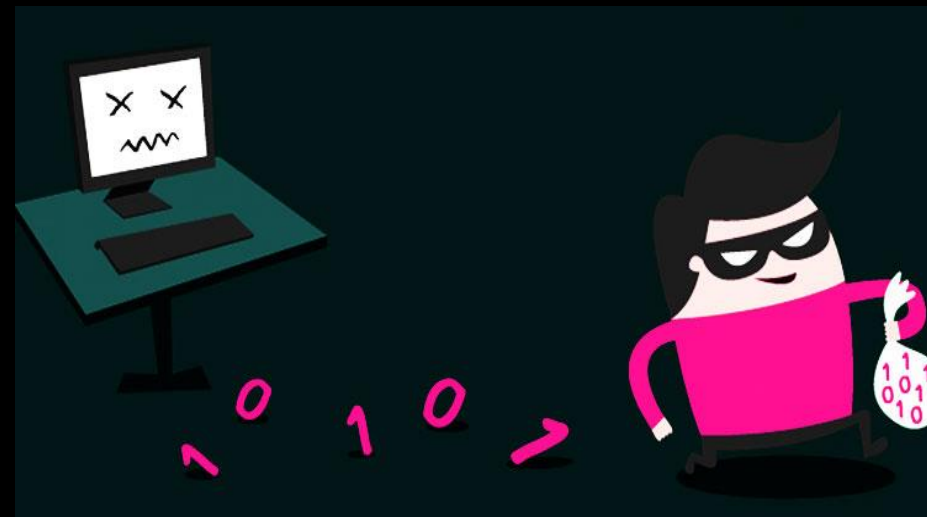
Given that Anna is age 15, COPPA is not a concern. For FERPA however, if Anna was enrolling in an online degree from University of Illinois, there may be certain obligations that we have in dealing with access to student records.





## Data Security Breach

Imagine Anna's information was part of an unauthorized data hack targeting EdTech companies. How should Coursera respond?



- ➔ Gather information and neutralize the situation
- ➔ Create a response team
- ➔ Communicate (Prepare PR response)
- ➔ Aftermath Handling



## How does Coursera deal with the various privacy concerns?

With many privacy concerns to monitor, our legal team works closely with various cross function teams:

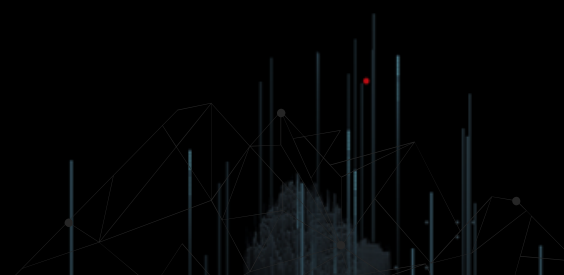


- Monitoring privacy laws and regulations in relevant jurisdictions and working on compliance filings
- Drafting privacy policies and contracts that protect our learners but allow flexibility for improving services and changing technology
- Providing avenues for learners to ask about their privacy concerns
- Establishing a data breach response plan
- Building privacy features into our product from ground up



## Questions to consider for your product on privacy

- What type of information are you collecting? (Is this PII?)
- Who are you collecting information from and what countries are they from?
- Have you received explicit consent from the data subject?
- What are the laws and regulations about privacy that apply to my product? (any local laws?) Have I built product features that comply with these laws?
- When a data breach happens, do I have a concrete, easy to operationalize plan?



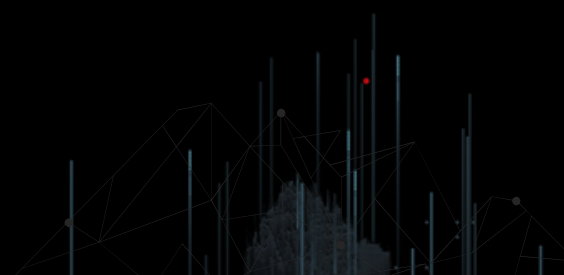


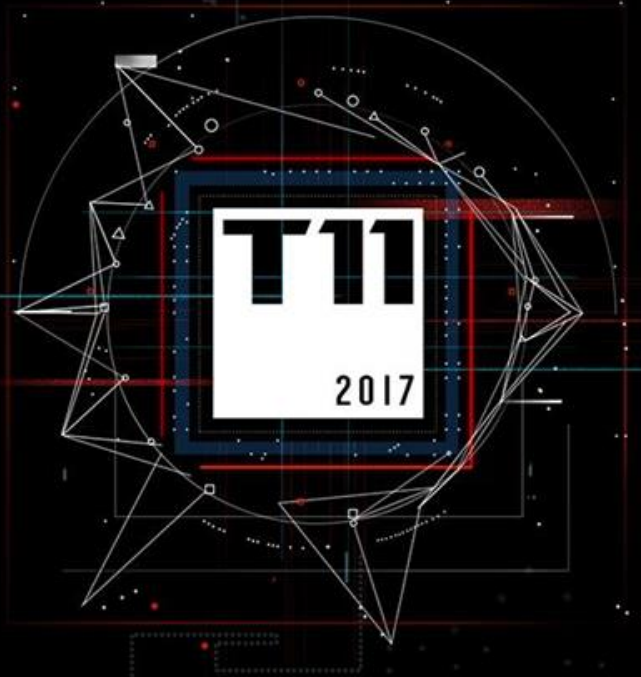
**Any Questions? Please reach out to us.**

**course**ra

email : [hyang@coursera.org](mailto:hyang@coursera.org)

email : [vinod@coursera.org](mailto:vinod@coursera.org)





**THANKS**