



Orientation to ISA5810

Yi-Shin Chen

Institute of Information Systems and Applications

Department of Computer Science

National Tsing Hua University

yishin@gmail.com

About Instructor

陳宜欣 Yi-Shin Chen

■ Currently

- Professor at NTHU CS
- Director of IDEA Lab

■ Education

- Ph.D. in Computer Science, USC, USA
- M.B.A. in Information Management, NCU, TW
- B.B.A. in Information Management, NCU, TW

■ Courses (*all in English*)

- Research and Presentation Skills
- Advanced Database Systems
- Data Mining: Concepts, Techniques, and Applications
- Data Structures
- Logical Thinking of Informatics



About Teaching Assistants



Didier Salazar



Retnani Latifah



Tadeus Tamus



Erdenejargal
Galtsuren



Tanisha Singh



Aliah Gie Zabala

About Teaching Assistants



Yung-Li Chan



Cheng-Yen Wu



Andy Chang



Chih-Yu Chuang



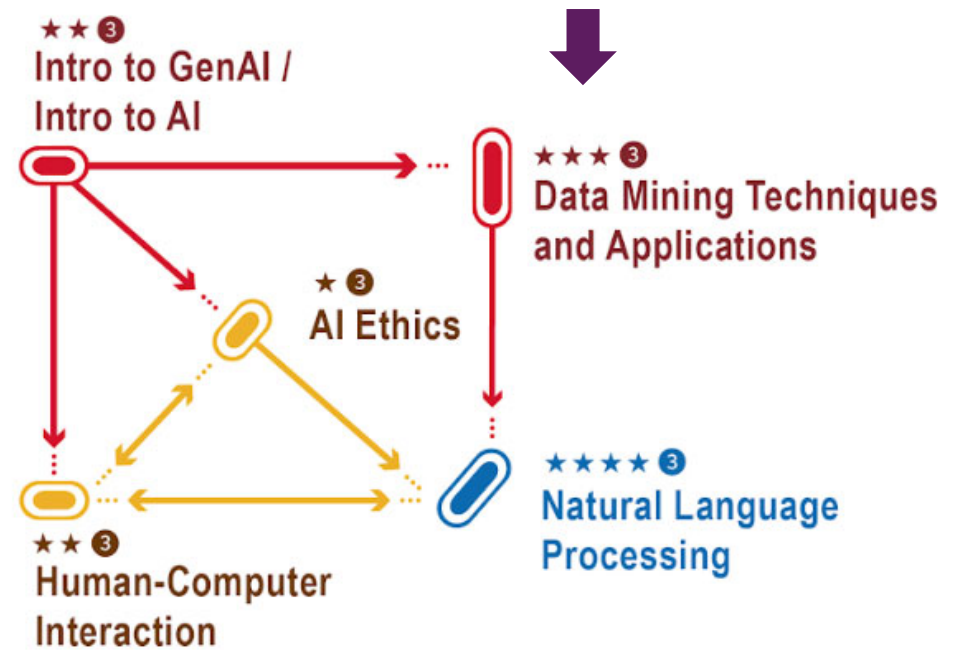
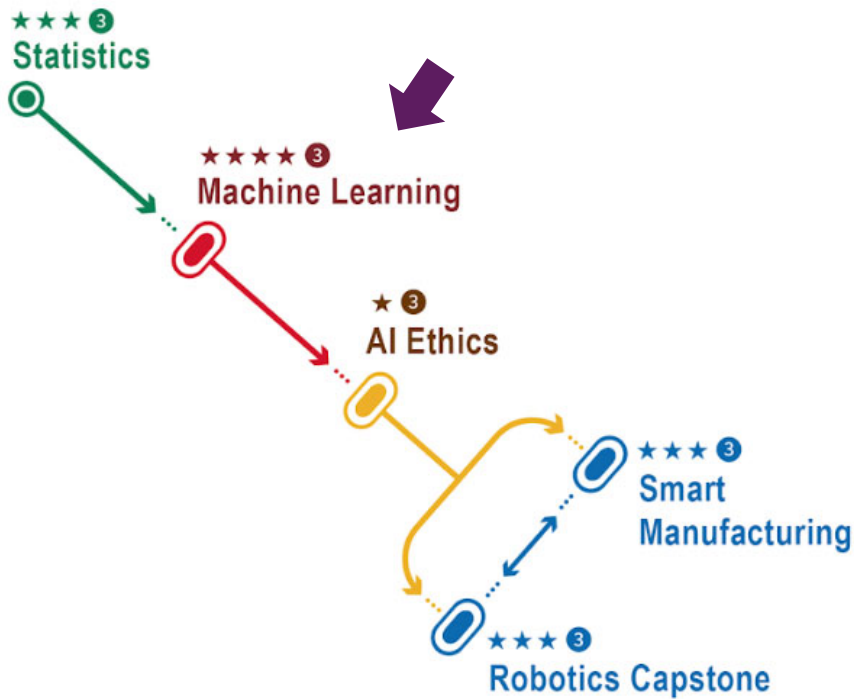
En-Chia (Andy) Hsu

Contact:
NTHU2024DM@gmail.com



114-1 Data Mining

■ Time: Monday 9:00~12:00



Reading Materials

■ Text Book:

- Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Addison Wesley

■ Additional Reading:

- Check course group website regularly

■ Handouts:

- Will be available at least one day before the class

General Information

■ Lectures:

- Mondays 9:00-11:40PM

■ The course web page:

- <https://www.cs.nthu.edu.tw/~yishin/courses/ISA5810/ISA5810-2025.html>

■ Class Platform:

■ Teams:

<https://teams.microsoft.com/l/team/19%3AAsOi679bnyEvWPbX4vQyKlXi9ndva2dN1NIBk9XqkvB81%40thread.tacv2/conversations?groupId=af7105ed-289a-4b7d-8c9e-b6089d8be293&tenantId=6c3bc511-43c7-4596-baeb-2335c69c41f1>

- YouTube: https://www.youtube.com/@NTHU_ISA5810_DataMining

- NTU Cool: <https://cool.ntu.edu.tw/courses/50367>

114-1 (2025 Fall)

- Home
- Syllabus
- Modules
- Announcements
- Assignments
- Quizzes
- Discussions
- Grades
- People
- Files

114-1_TAICA_Data Mining : Concepts, Techniques, and Applications 資料探勘與應用_主導課程



About


Data mining serves as a crucial field that leverages advanced algorithms to reveal hidden, yet invaluable insights buried within extensive datasets. These algorithms are drawn from a multitude of areas such as machine learning, artificial intelligence, pattern recognition, statistics, and database systems, working together to facilitate a deeper understanding and analysis of data.

This course, ISA5810: Data Mining: Concepts, Techniques, and Applications, is designed to equip you with the foundational knowledge and hands-on experience needed to delve into the expansive world of data mining. Whether you are looking to enhance your skill set or embark on a new career path, this course will serve as a stepping stone to achieving your goals.

The curriculum encompasses a range of topics that will introduce you to the core concepts and techniques

Activities

- Several in-class and after-class activities
 - Including orientation quiz (9/1~9/22)
- Two lab assignments
- One short presentations
- One project for analyzing data
- One closed-book final exam
 - You can bring one A4 page paper

⋮	Activities	✓	⋮
⋮	 Orientation Quiz 99 pts	✓	⋮

Grading Policies

■ Policy:

- Two assignments: 20%
- One short presentation: 10%
- One project: 25%
- One exam: 35%
- Class participation: 10% (In-class or after-class)

Offered in English

■ Why? The school asked for such courses

■ Obstacles:

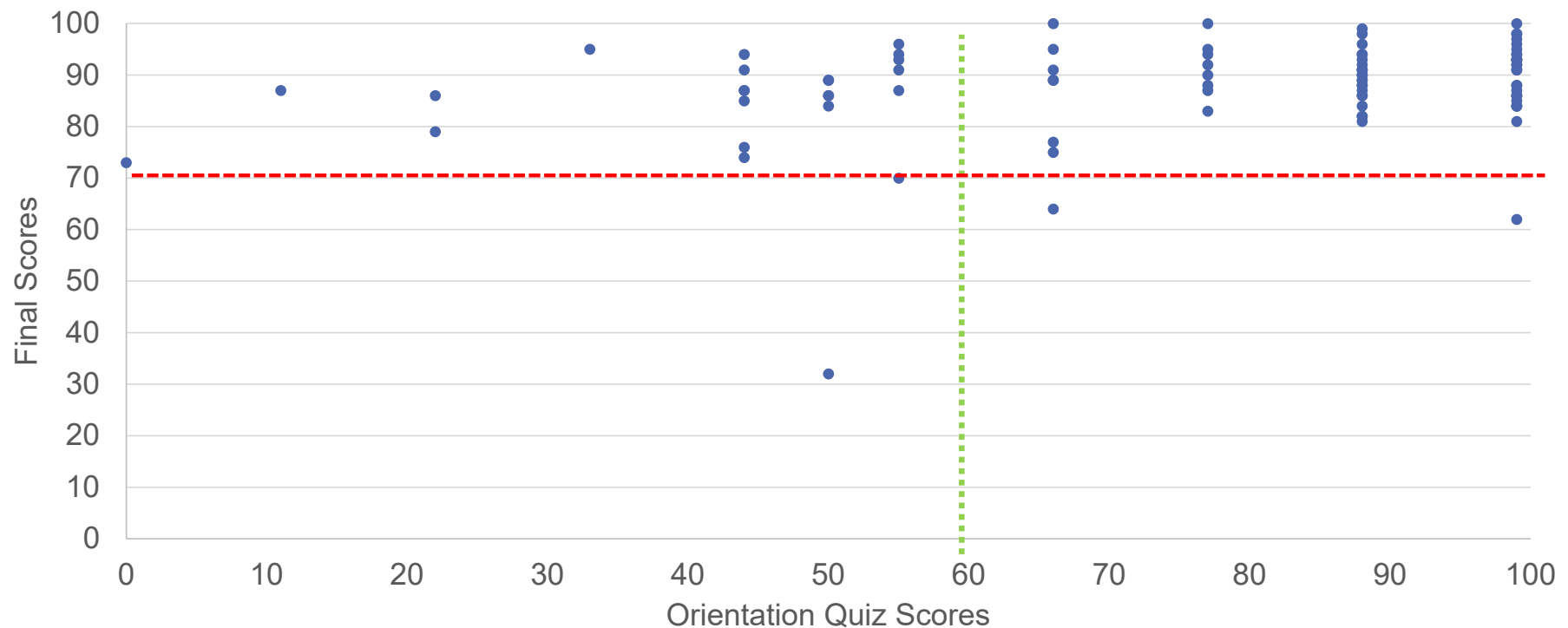
- Do you believe in yourself?
- Can you work hard?
- Can WE work together?

■ Why the orientation quiz? (After-class)

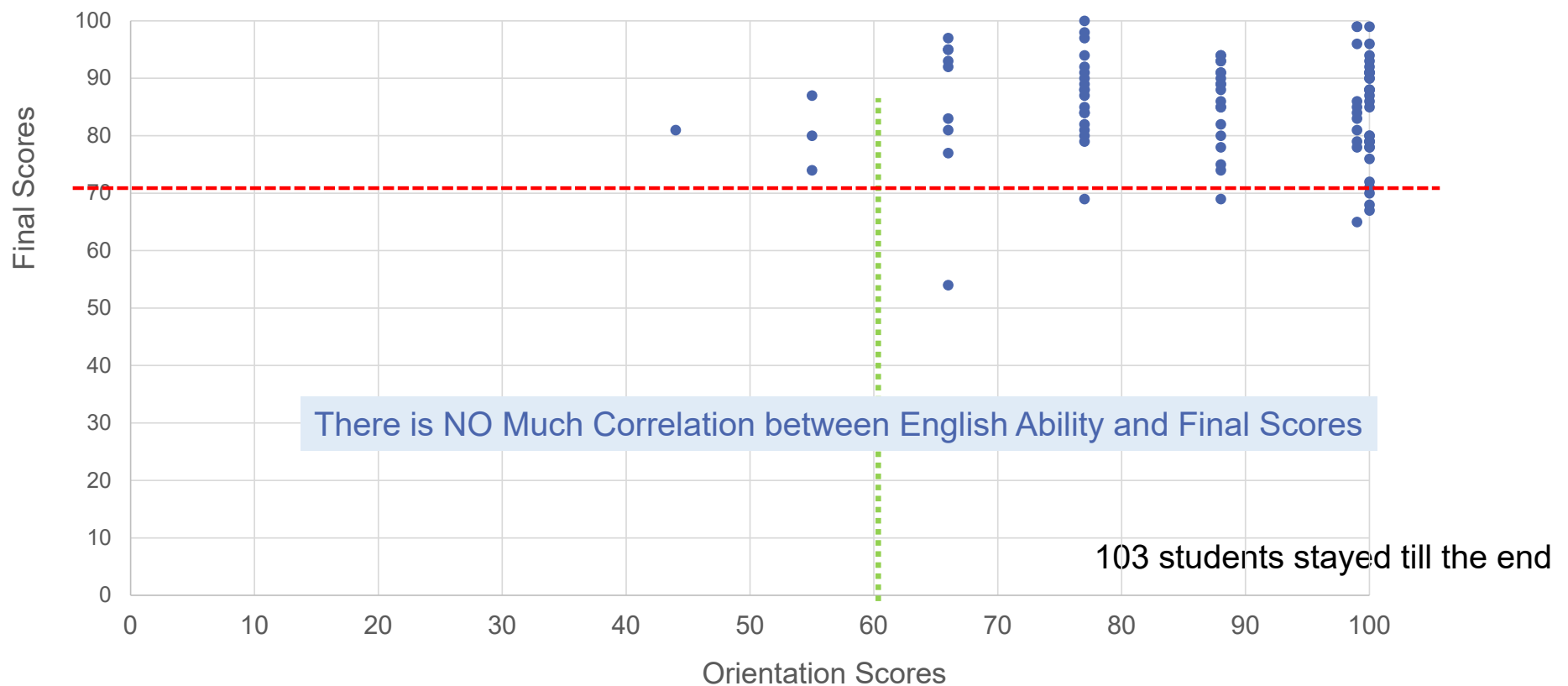
- In NTU Cool
- To understand the goal of the course
- To understand yourself
- To know my accent (Can you get used to it?)
- To know how much we can improve

Orientation Quiz vs. Final Grades (Data Mining, 2021)

There is NO Much Correlation between English Ability and Final Scores



Orientation Quiz vs. Final Grades (Data Mining, 2023)



Letter-based Grades

- Grading adjustments are made according to the respective grading distribution
- Each grading scheme will adhere to its respective grading distribution curve
- Here's the modified breakdown:
 - A+: Within the top 5% and scoring above 90
 - A: Within the top 15%
 - A-: Within the top 25%
 - B+: Within the top 40%
 - B: Within the top 55% and scoring above 75
 - B-: Initial grade exceeds 70
 - C/D/F: Noteworthy Lower Percentile

Course Academic Integrity Policy

- Projects should be written without referencing others' codes
- You are not allowed:
 - Copying all or part of someone else's work (including generated AI techniques, such as ChatGPT)
 - Giving another student in the class a copy of your work
 - Consulting with others (including generated AI techniques) during an exam
- Students who violate this policy
 - In assignments: **negative credit**
 - In the exams: **final score will be ZERO**

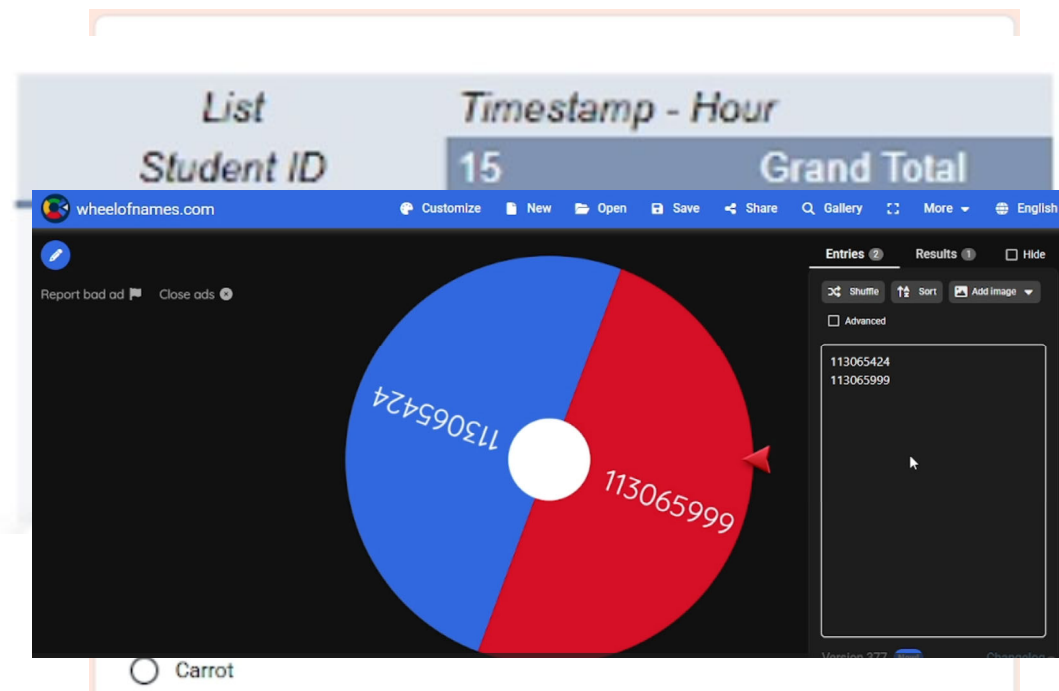


Participation



Encourage To Stay in Class

- Lottery for Colab Compute Units
 - 100 Compute Units
 - 500 Compute Units
- Held in Each Lecture & Final Demo
 - For students attending in the classroom
 - For students joining online in real time





Schedule



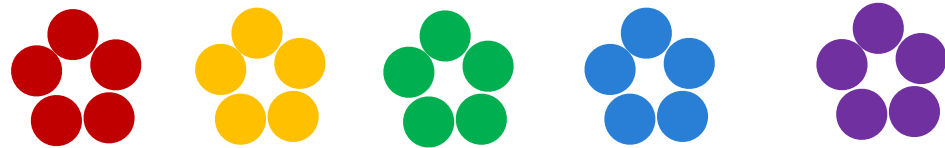
Schedule (16 Weeks and Distant Learning in Some Sessions)

Date	Topics
9/1	Introduction and overview
9/8, 9/15	Overview and Data
9/15 (Recording)	Lab 1
9/22, 9/29 (Recording)	Classification
10/13	Text Mining & Project Progress Report
10/20	Lab 2
10/27, 11/3	Text Mining

Schedule (Contd.)

Date	Topics
11/10	DM Clustering Part1 & Project Progress Report
11/17	DM Clustering Part 2
11/24	Association Rules
12/1	Student Paper Presentation (NTHU students should be in classrooms)
12/8	Final Exam
12/16	Final Demo

Presentation Phases

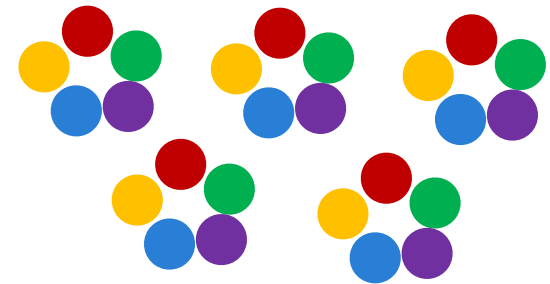


■Phase 1:

- Students presenting the same section will be grouped together.
- Within these groups, participants should discuss and clarify their understanding of the section.

■Phase 2:

- Students will regroup with new members.
- Begin by answering the questionnaire related to Phase 1
- Each student will present their designated section



■Post Phase 2 Activities:

- Go back to the class
- Complete the Phase 2 questionnaire
- Fill in one-minute reflection

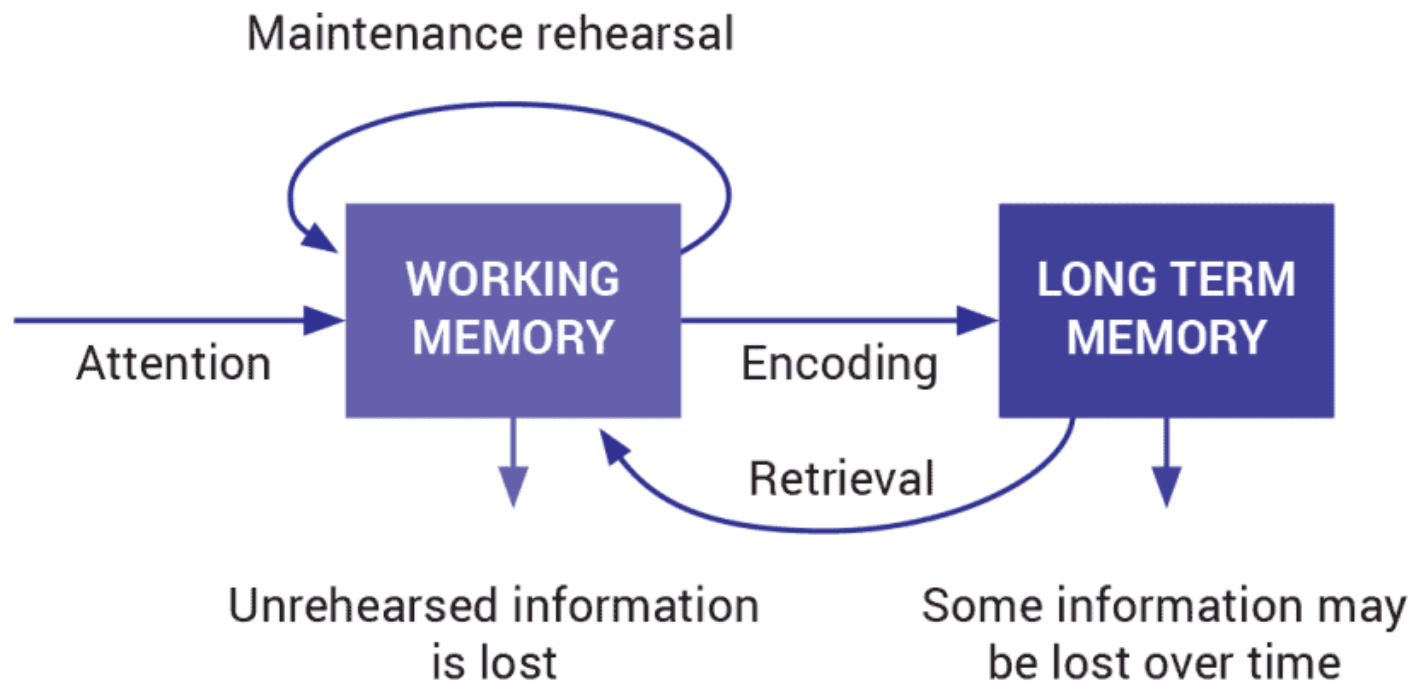




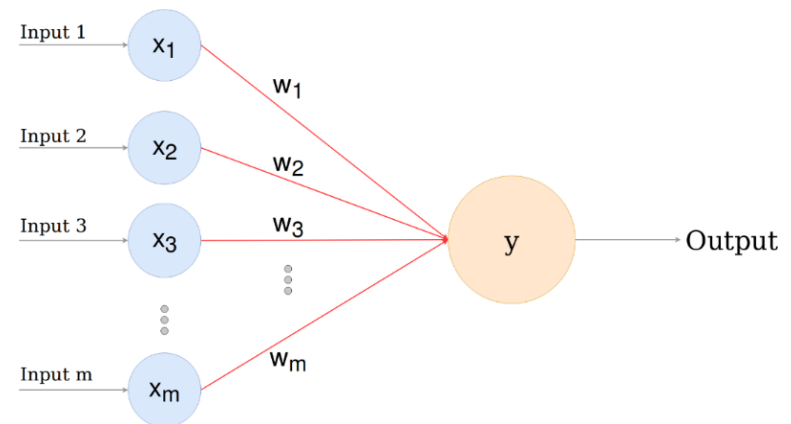
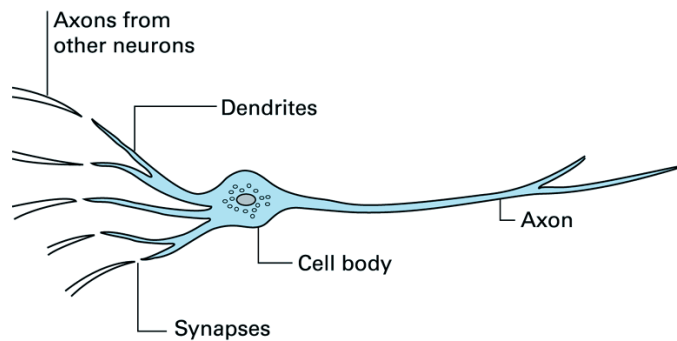
Effective Learning in EMI



Cognitive Load Theory



Neural Networks - Perceptron (1958)



Unknow (Problem)

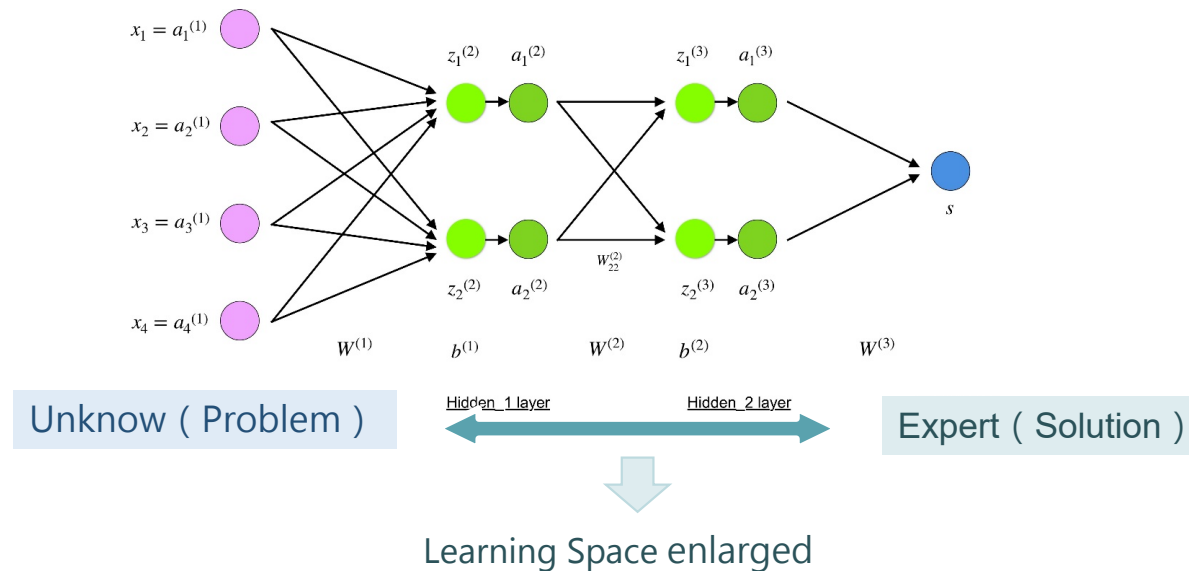
Expert (Solution)



Dr. Becky Kennedy: Learning Space

- Network is programmed by adjusting weights using feedback from examples
- Cannot solve XOR

Backpropagation (1986)



- Adding hidden layers
- Mistakes backpropagate

backpropagate → AI: Electricity

backpropagate → Human: Resilience

Deep Learning

Neural Networks
in 1990



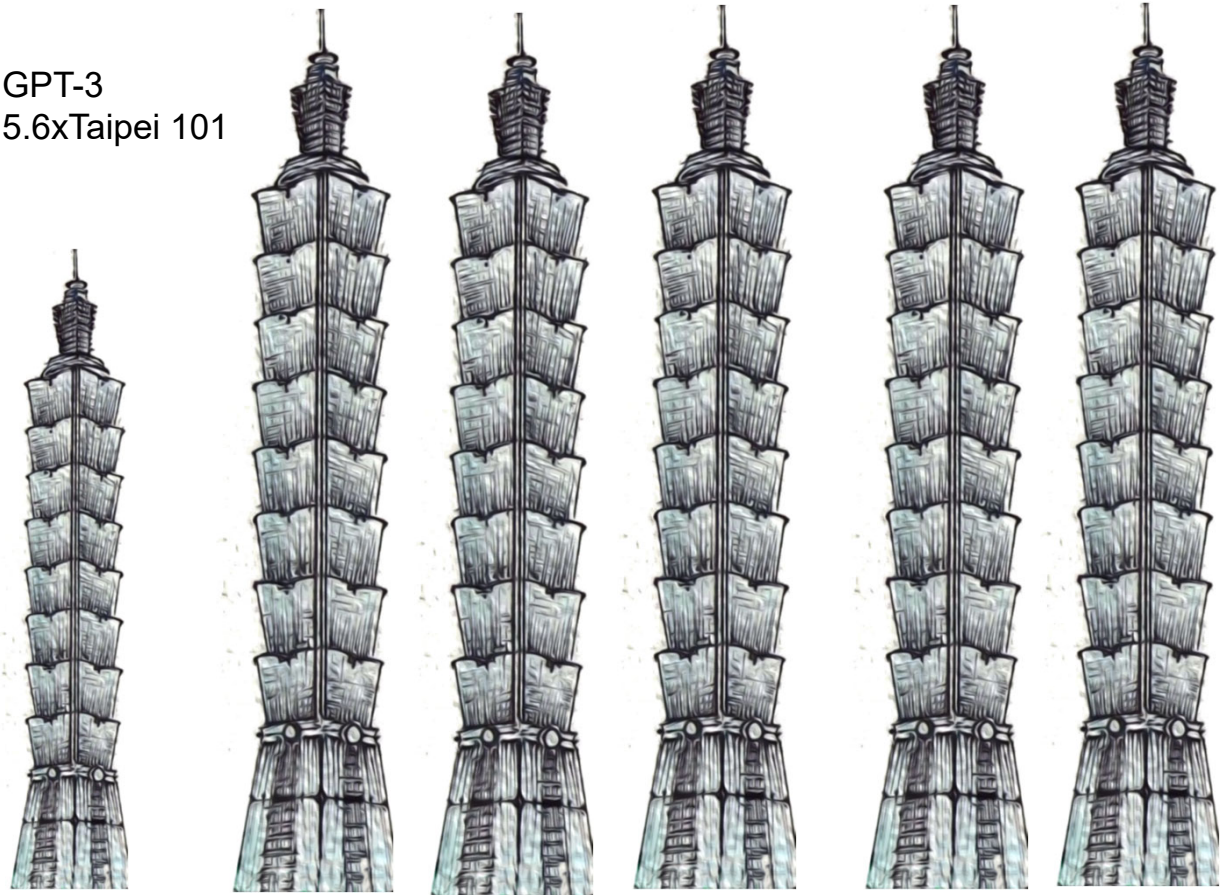
3mm

LLM in 2016



20cm

GPT-3
5.6xTaipei 101



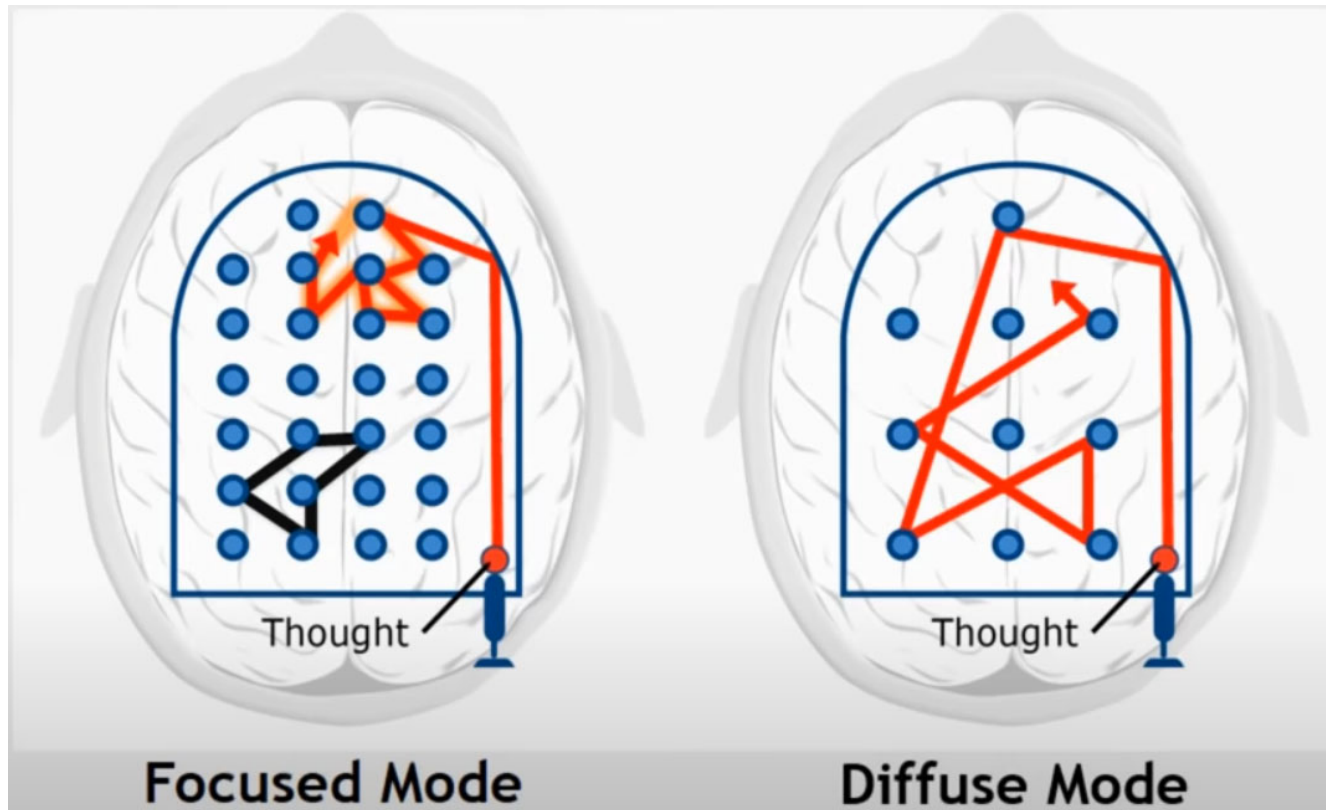
Resilience

- From Dr. Becky Kennedy
- Helping our kids be good at struggling
 - is a guiding principle in raising
 - not only resilient kids
 - but antifragile kids
 - and really confident, capable kids

Learning can be frustrating, but the joy comes after you have learned.

Happiness depends on learning, but surprisingly, it doesn't depend on reward (from Prof. Bastien Blain)

Focused Attention vs. Diffuse Mode



Behavior First

- Ordinary thought

- Sensation → perception → feeling → thought → behavior

- To control our nervous system, run it backward

- Behavior → thought → feeling → perception → sensation

- Problem Solving Phases

- Define the problem
- Problem analysis
- Generate possible plans/solutions

Focus on what you can accomplish before giving in to panic

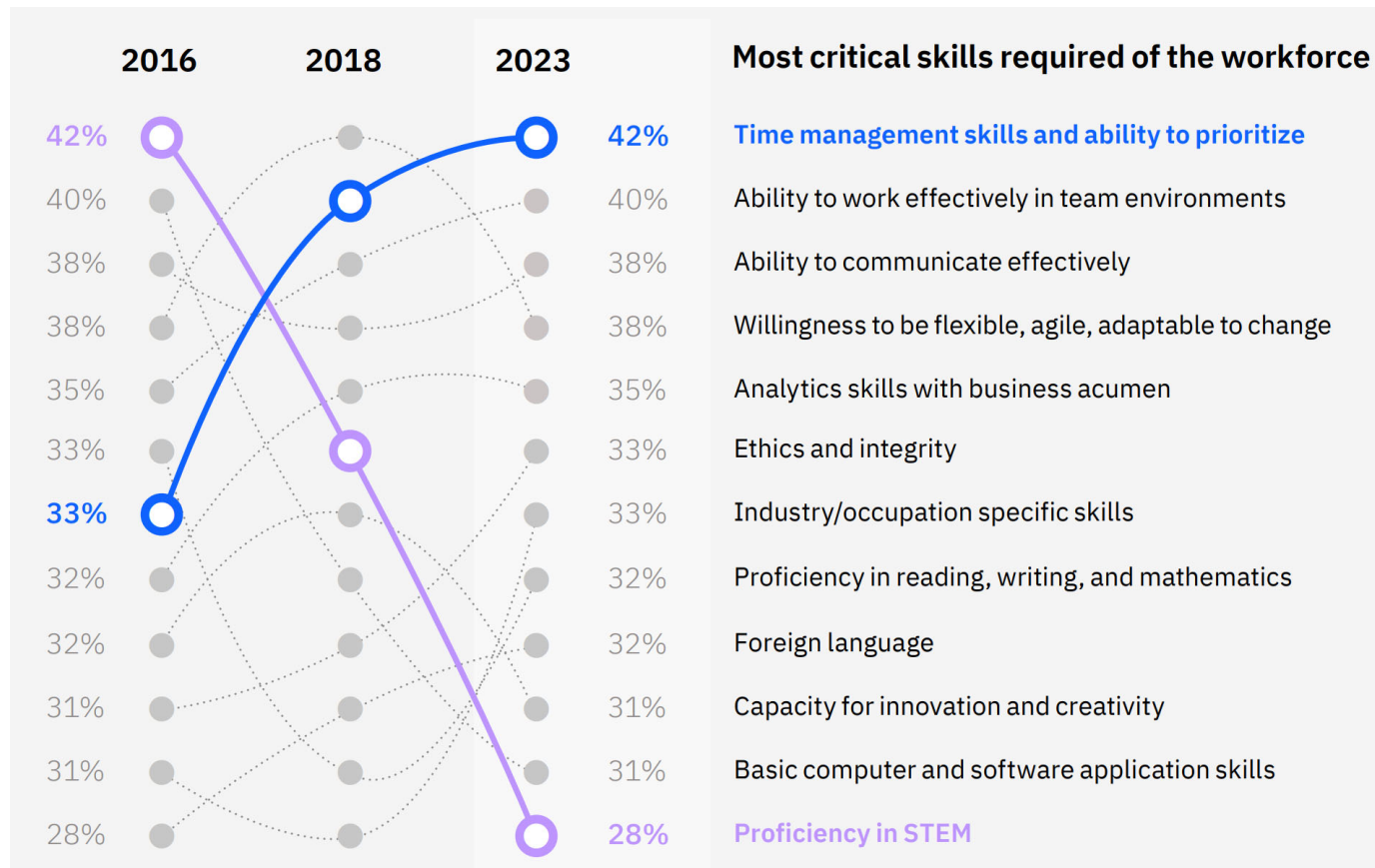
Quotes

- “The capacity to learn is a gift; the ability to learn is a skill; the willingness to learn is a choice.” – Brian Herbert
- “Learning is not the product of teaching. Learning is the product of the activity of learners.” – John Holt



People Skills Matter

IBM, Augmented work for an automated, AI-driven world,
<https://www.ibm.com/downloads/cas/NGAWMXAK>



Introduction to Data Mining @ Yi-Shin Chen

Activities in Classes

- Some lectures
- Preview activities
- Many class activities
- One-minute personal reflections