Human-Centered Artificial Intelligence

Lecture 2: AI types and capabilities

https://chatw.ch/hcai25

Chat Wacharamanotham

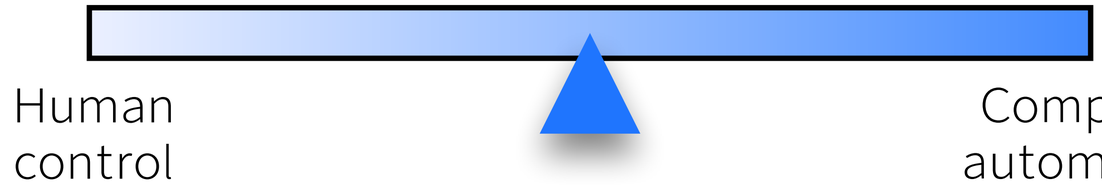
Fall Semester 2025

Intended learning outcomes

By the end of this course, students will be able to

- ❏ Analyze user tasks in terms of the levels of control and automation.
- ❏ Recognize that different types of AI systems have different capabilities.
- ❏ Motivate the importance of reliability, safety, and trustworthy AI systems.
- ❏ Choose types of AI that are suitable for the user tasks in their design project.

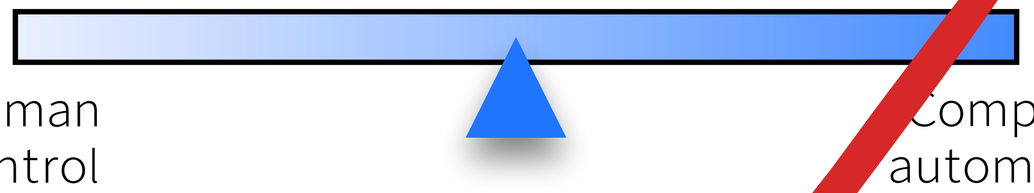
Shneiderman’s Control × Automation quadrants



Sheridan & Verplank's levels of autonomy	**The Computer:**
10 (High)	decides everything and acts autonomously, ignoring the human
9	informs the human only if the computer decides to
8	informs the human only if asked
7	executes automatically, then necessarily informs the human
6	allows the human a restricted time to veto before automatic execution
5	executes the suggestion, if the human approves
4	suggests one alternative
3	narrows the selection down to a few
2	offers a complete set of decision-and-action alternatives
1 (Low)	offers no assistance; the human must take all decisions and actions

Shneiderman (2022) Human-Centered Artificial Intelligence. Oxford University Press. (Chapters 6 and 8)

Misleading one-dimensional thinking



Human control

Computer automation

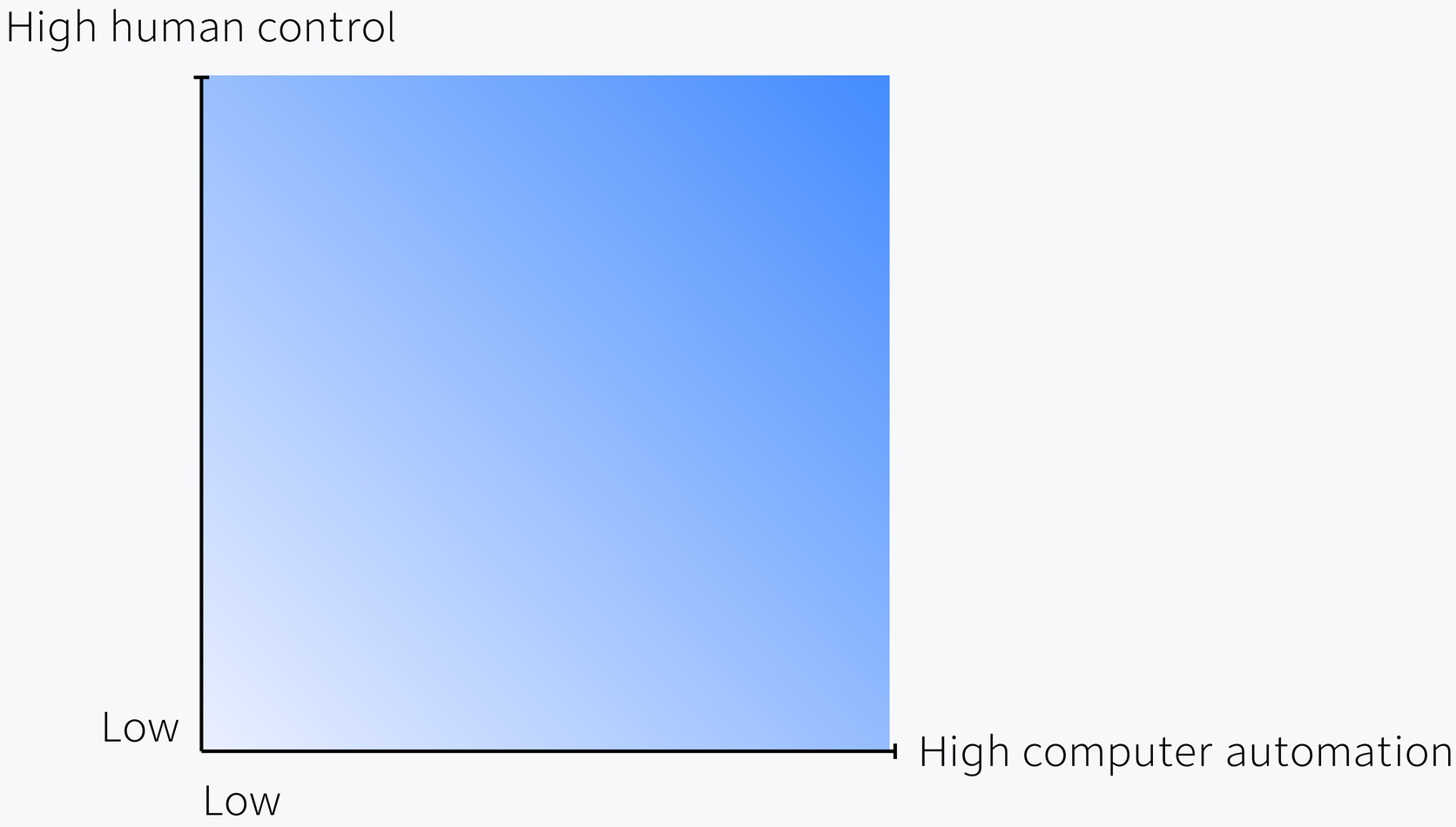
Sheridan & Verplank's levels of autonomy

	The Computer:
10 (High)	decides everything and acts autonomously, ignoring the human
9	informs the human only if the computer decides to
8	
7	
6	tic execution
5	executes the suggestion, if the human approves
4	suggests one alternative
3	narrows the selection down to a few
2	offers a complete set of decision-and-action alternatives
1 (Low)	offers no assistance; the human must take all decisions and actions

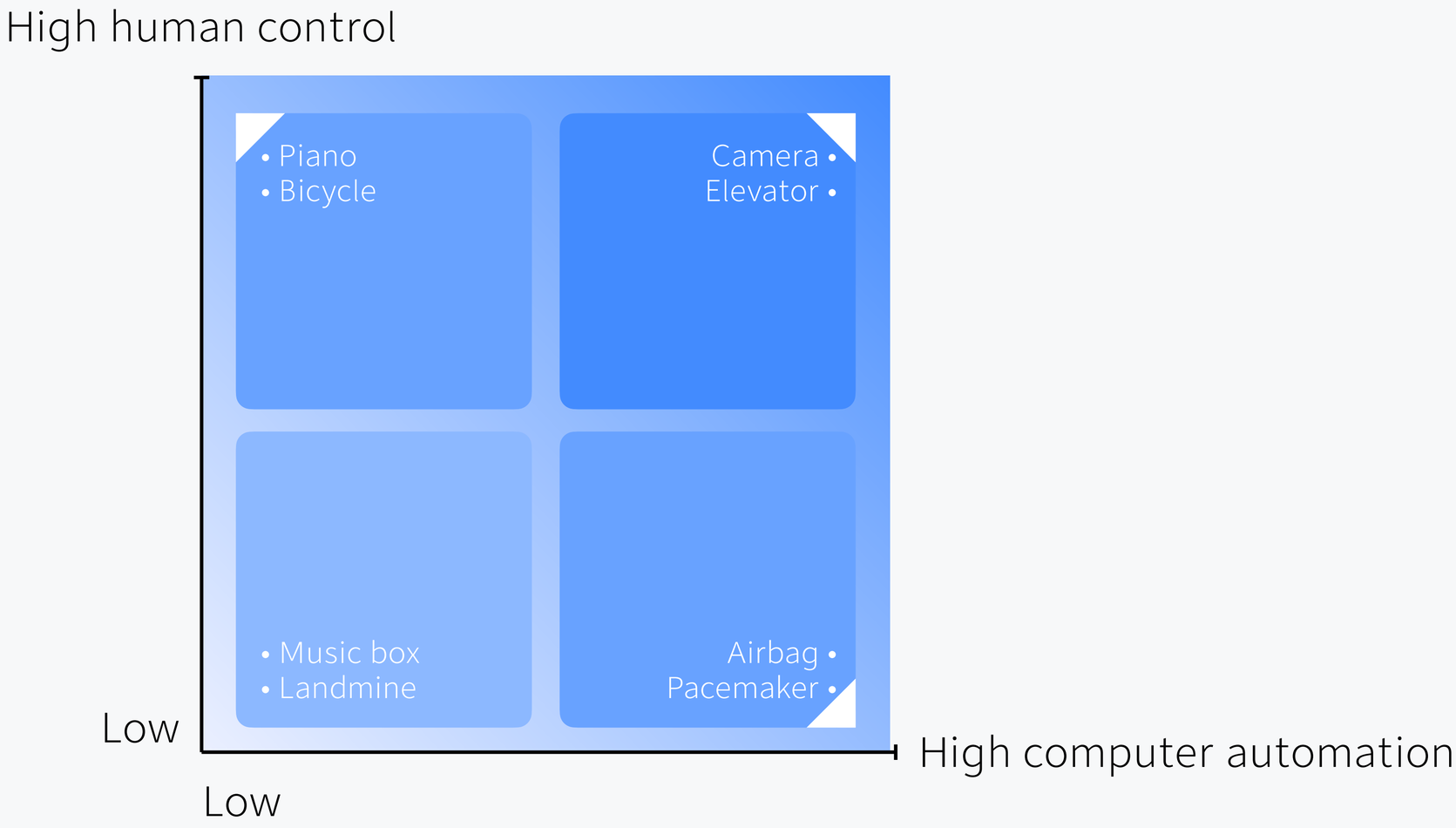
Control vs. automation trade-off is misleading!

Shneiderman (2022) Human-Centered Artificial Intelligence. Oxford University Press. (Chapters 6 and 8)

Shneiderman's Control × Automation quadrants

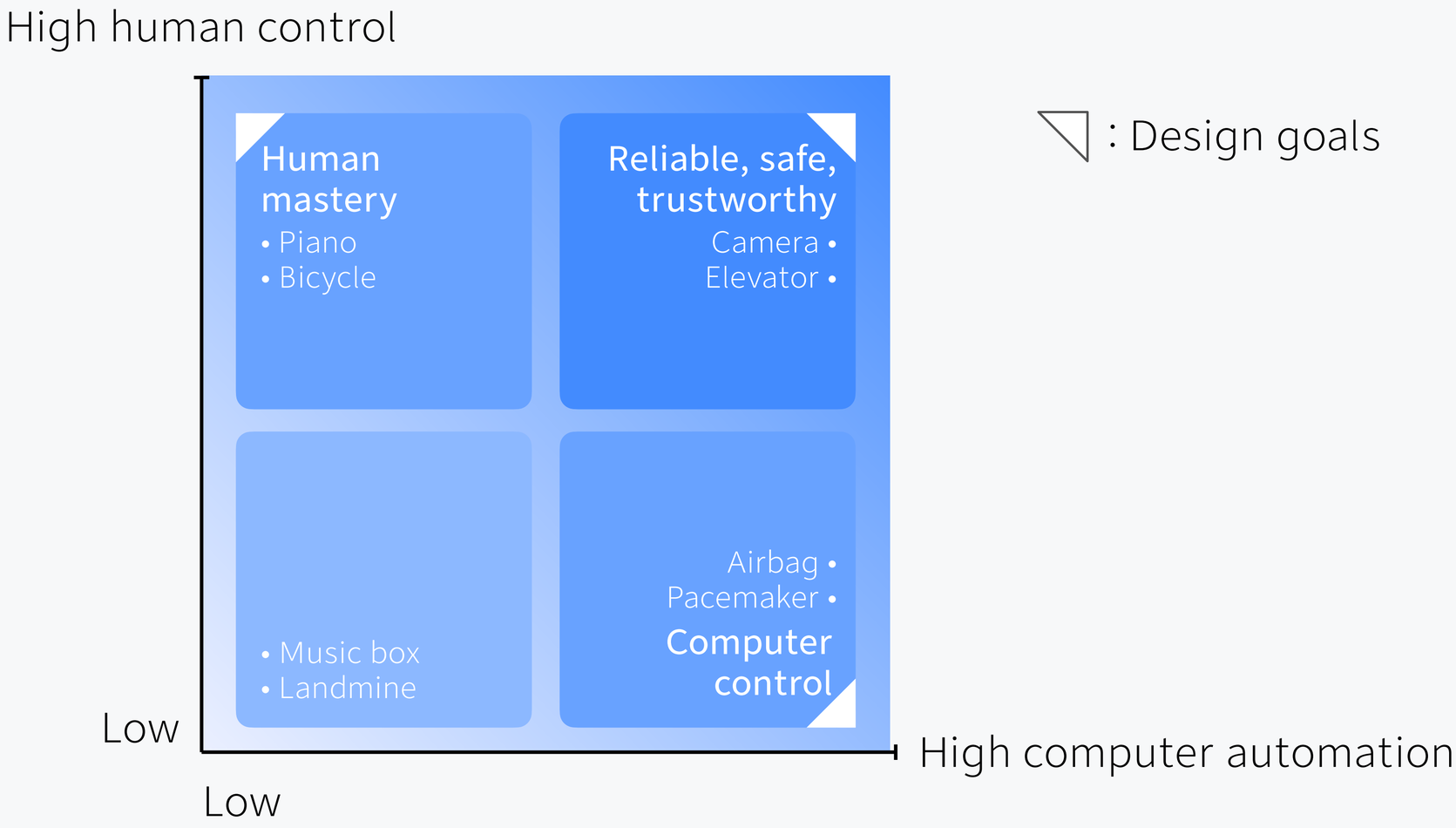


Shneiderman's Control × Automation quadrants



Shneiderman (2022) Human-Centered Artificial Intelligence. Oxford University Press. (Chapter 8)

Shneiderman’s Control × Automation quadrants

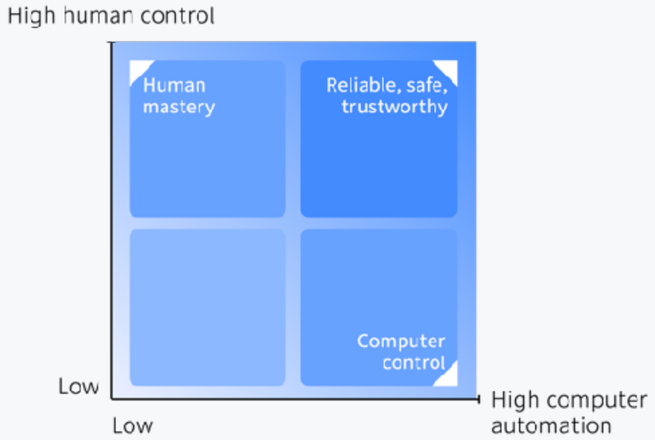


Shneiderman (2022) Human-Centered Artificial Intelligence. Oxford University Press. (Chapter 8)

Exercise: Control × Automation

⏳ 15 minutes 👥 work in pairs on one laptop

1. Go to the Miro board, find a place for your pair
2. Copy your list from the spreadsheet and paste onto the Miro board (with Ctrl+V or Cmd+V) as sticky notes
3. Work together to **place them in the control × automation quadrants**
   - Discuss **why you decided** on such a location
   - If you cannot agree on one location, make a copy, and place each copy in different locations
   - Aim for 10–12 sticky notes in the quadrants
4. Last 5 minutes: Reflect on what you learned from the discussion. Write 1–3 insights from this activity in the poll linked to the left of your diagram



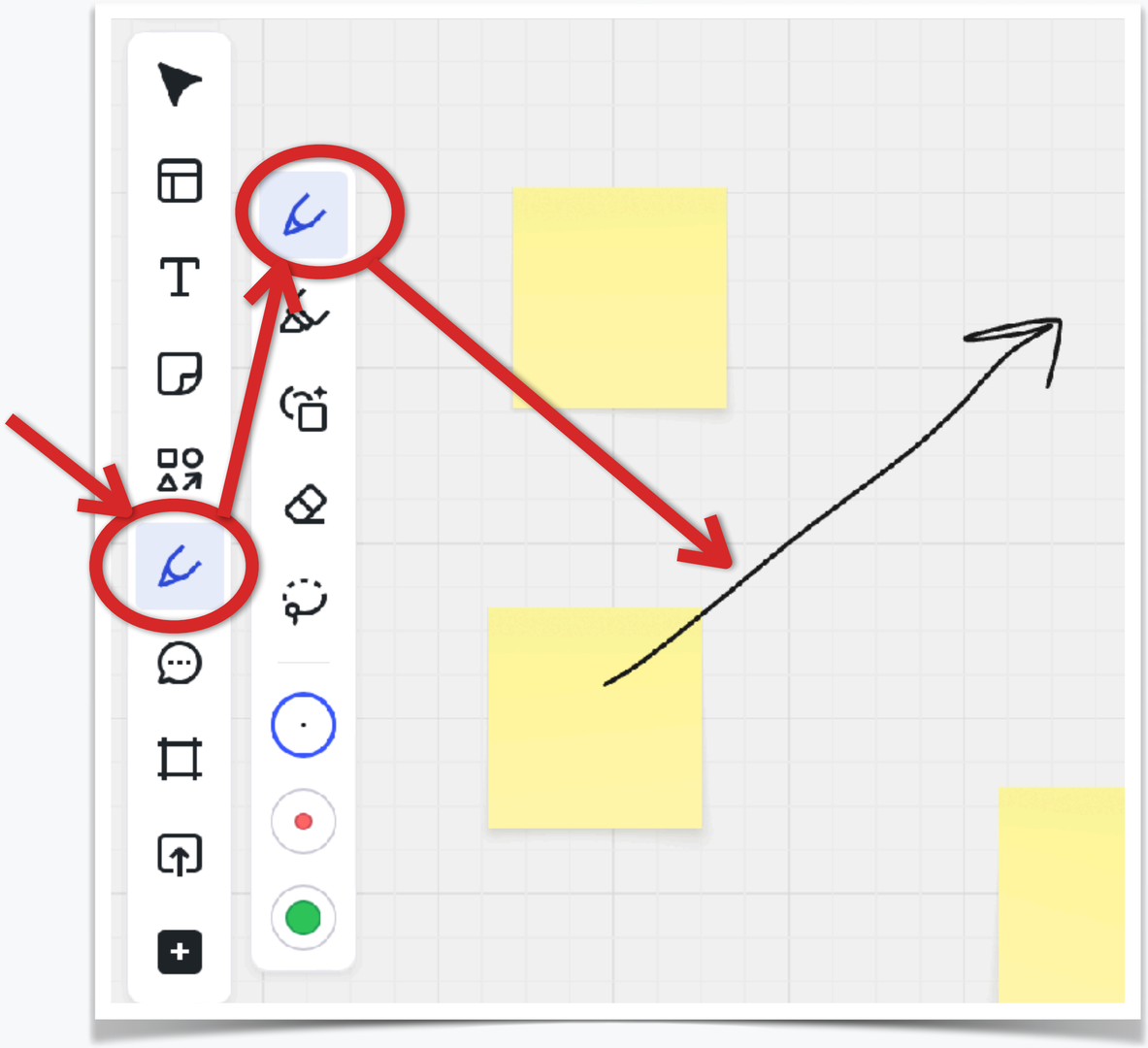
https://chatw.ch/hcai25

⏳ 10 minutes 👤👤work in pairs

Imagine a far future where technology has greatly advanced to an ideal state, and determine the location of each note.

- If it should be changed, use the pen tool to draw an arrow to the new location
- If not, you don’t need to do anything

Last 3 minutes: Write one insight on the poll



Evolved expectation

High human control

Human mastery

Reliable, safe, trustworthy

Computer control

Low

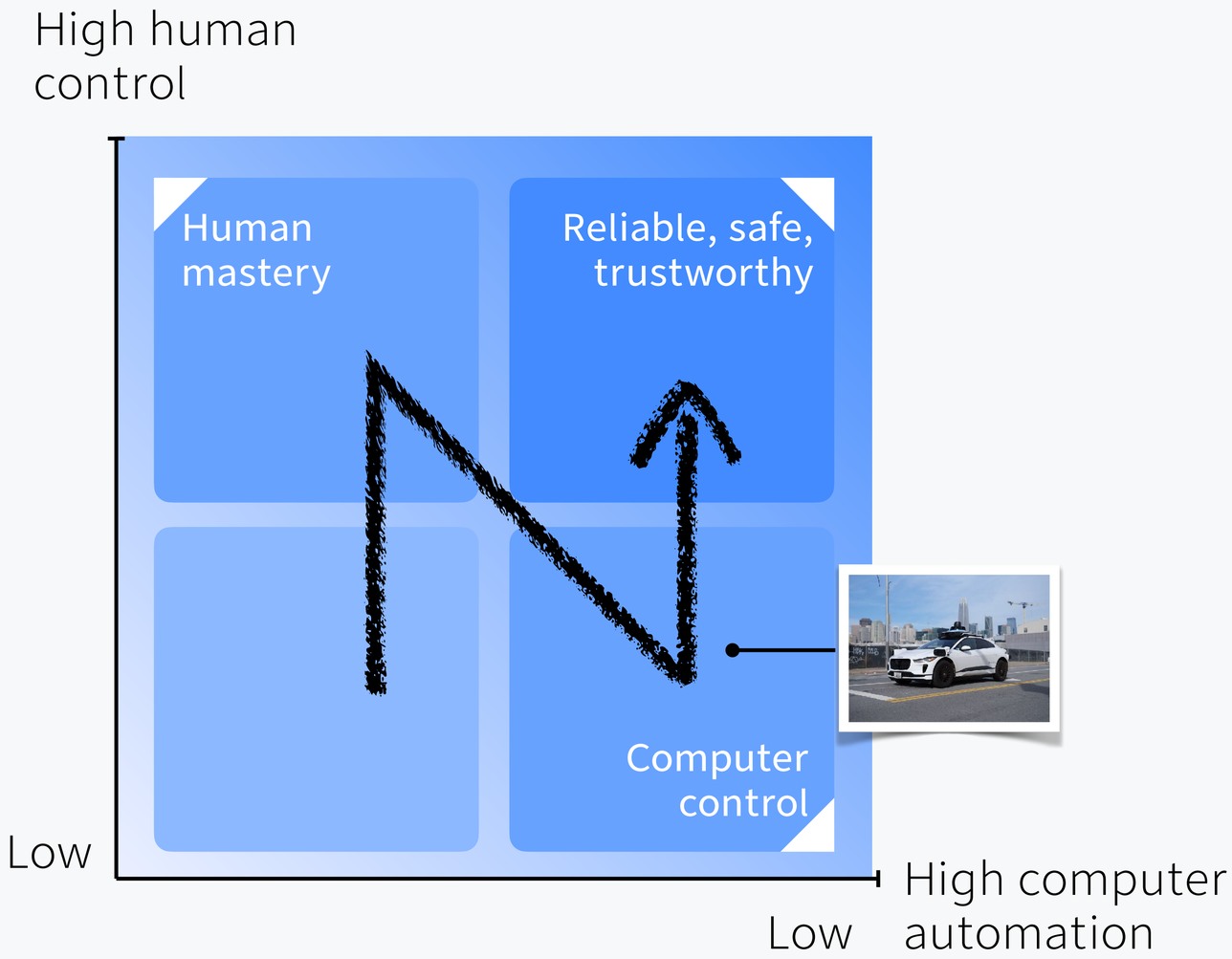
Low

High computer automation

When technology and regulation mature, we should expect better technologies

Images: Old car: Pixabay via Picryl • Cadillac by That Hartford Guy, via Wikipedia • Waymo car by DIlu via Wikipedia

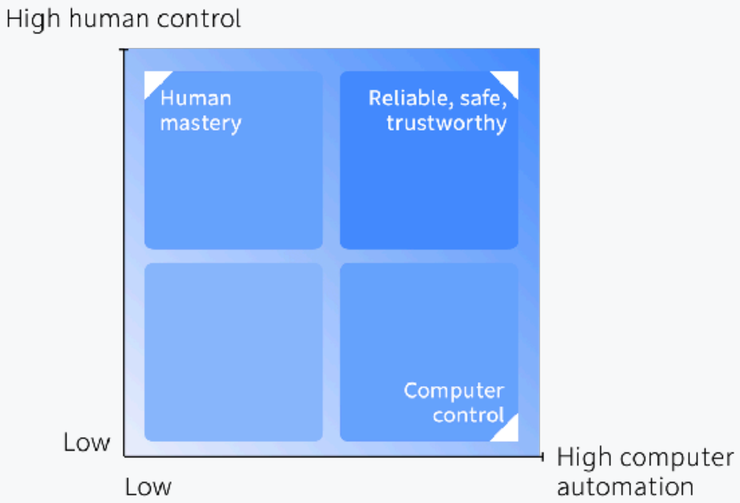
Human vigilance



- Computer controls become more reliable
- Human operators only rarely need to intervene
- Human attention lapse
- They can no longer intervene quickly and correctly

“sleeping at the wheel”

Images: Old car: Pixabay via Picryl • Cadillac by That Hartford Guy, via Wikipedia • Waymo car by DIlu via Wikipedia



"regular, honest, and cooperative behavior, based on commonly shared norms"

— Francis Fukuyama's definition of *social trust*

Reliable systems produce expected responses when needed. Supports human responsibility, fairness, and explainability.

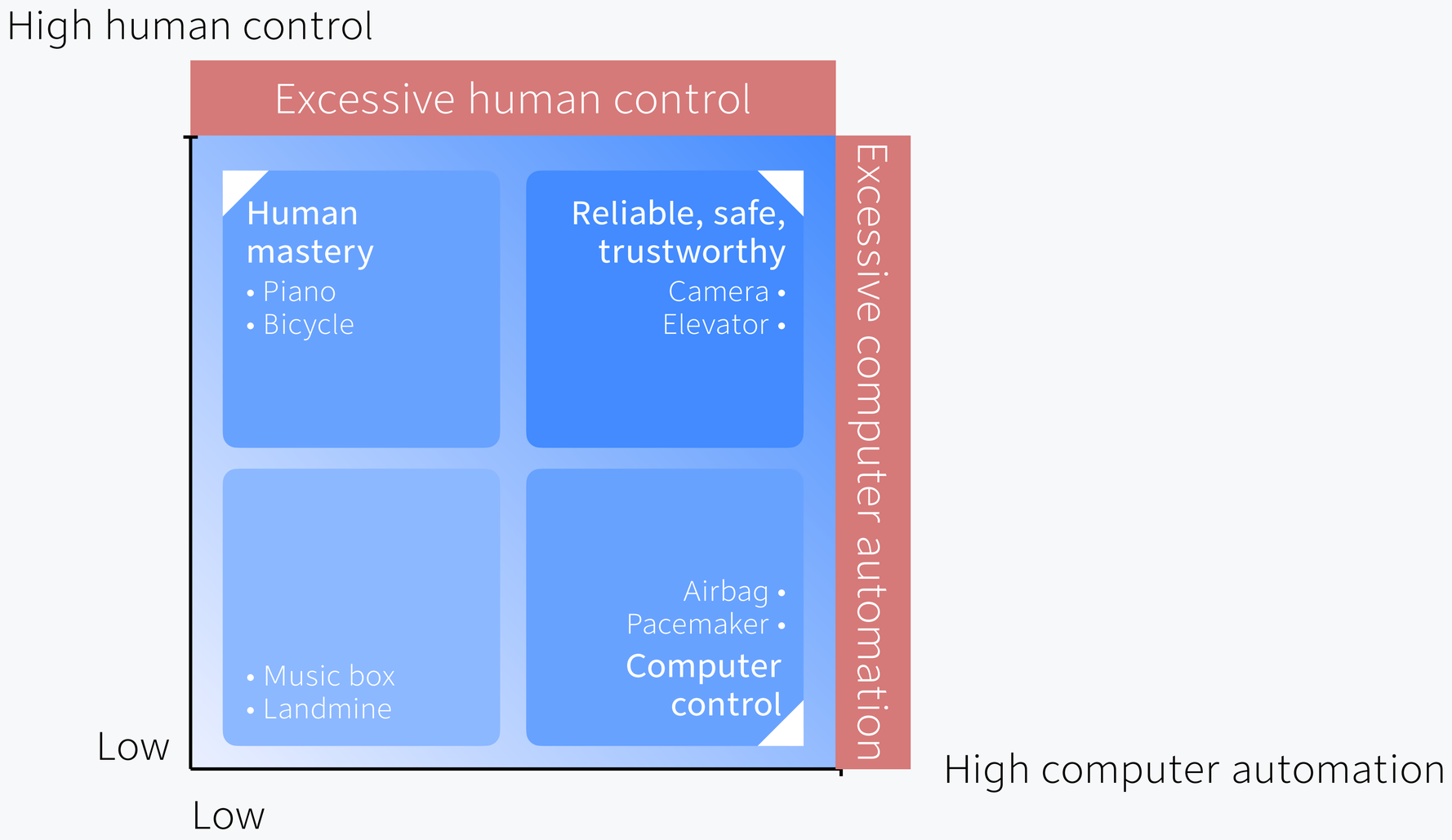
Cultures of safety are created by **managers** who focus on strategies that guide continuous refinement of training, operational practices, and root-cause failure analyses.

A **trustworthy system** is one that deserves trust, even though stakeholders struggle to measure

- **Consumers** do not have the skill or effort to assess it.
- Rely on **established independent organizations**, e.g., consumer advocacy group or respected auditing firms
- Rely on regulations by **governmental bodies**

 Shneiderman (2022) Human-Centered Artificial Intelligence. Oxford University Press. (Chapter 7)

Danger areas



 Shneiderman (2022) Human-Centered Artificial Intelligence. Oxford University Press. (Chapter 8)

Boeing 737 MAX's MCAS system

- Angle of attack sensor fails
- Automated system repeatedly pull the plane down
- Pilot had to repeatedly the plane pull up
- 346 people died from two crashes in 2018–19,

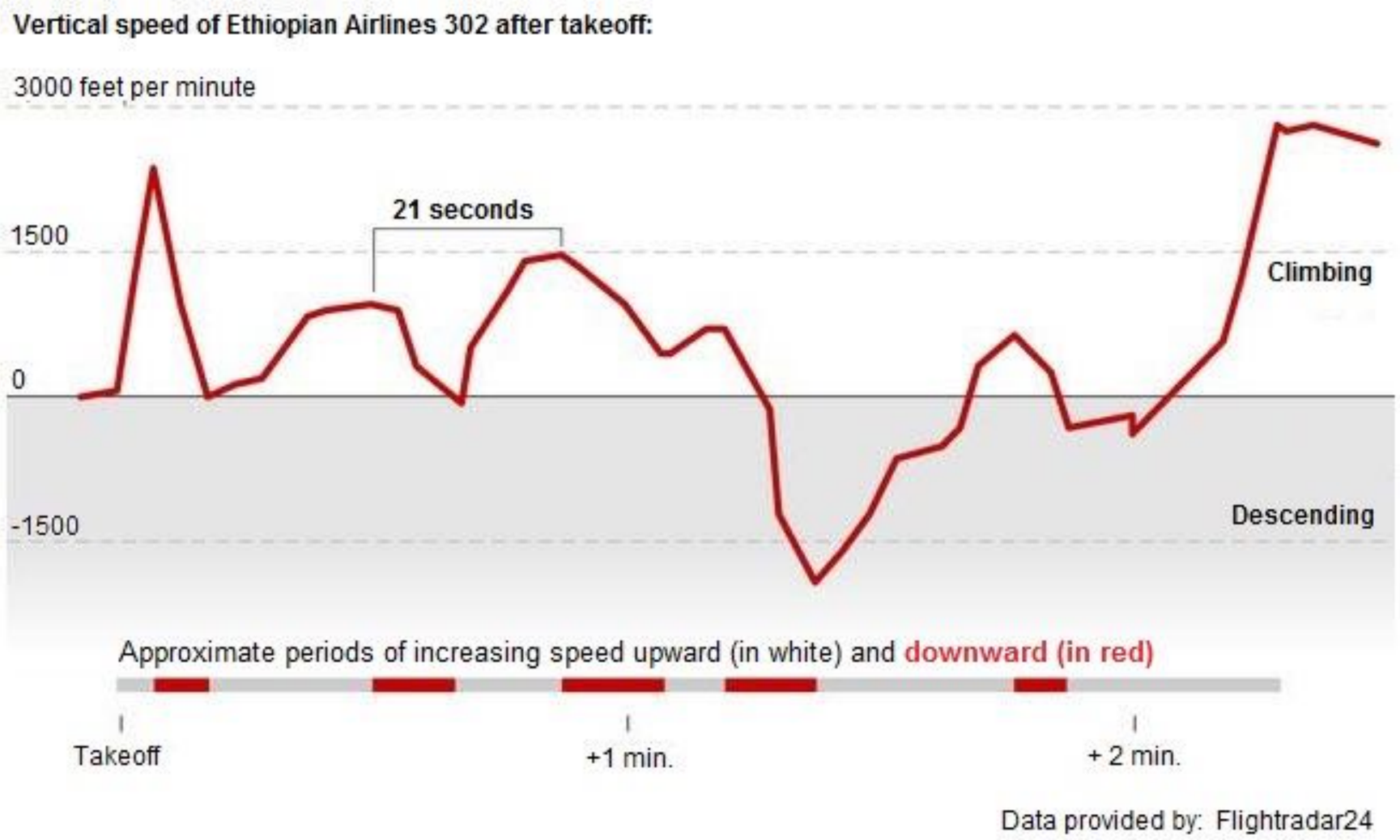


 Diagram: Zerohedge

Excessive human control example

Abbott pain-killer pump

- Nurse programs the concentration
  - intended: 5 mg / mL
  - entered: 0.5 mg / mL

- Patient received 10x too much!

- Device: ⬆, ⬇ buttons instead of number pad
  - Cost-saving vs. life-saving

Lifecare PCS Plus II Infuser type 4100

 Photo via eBay

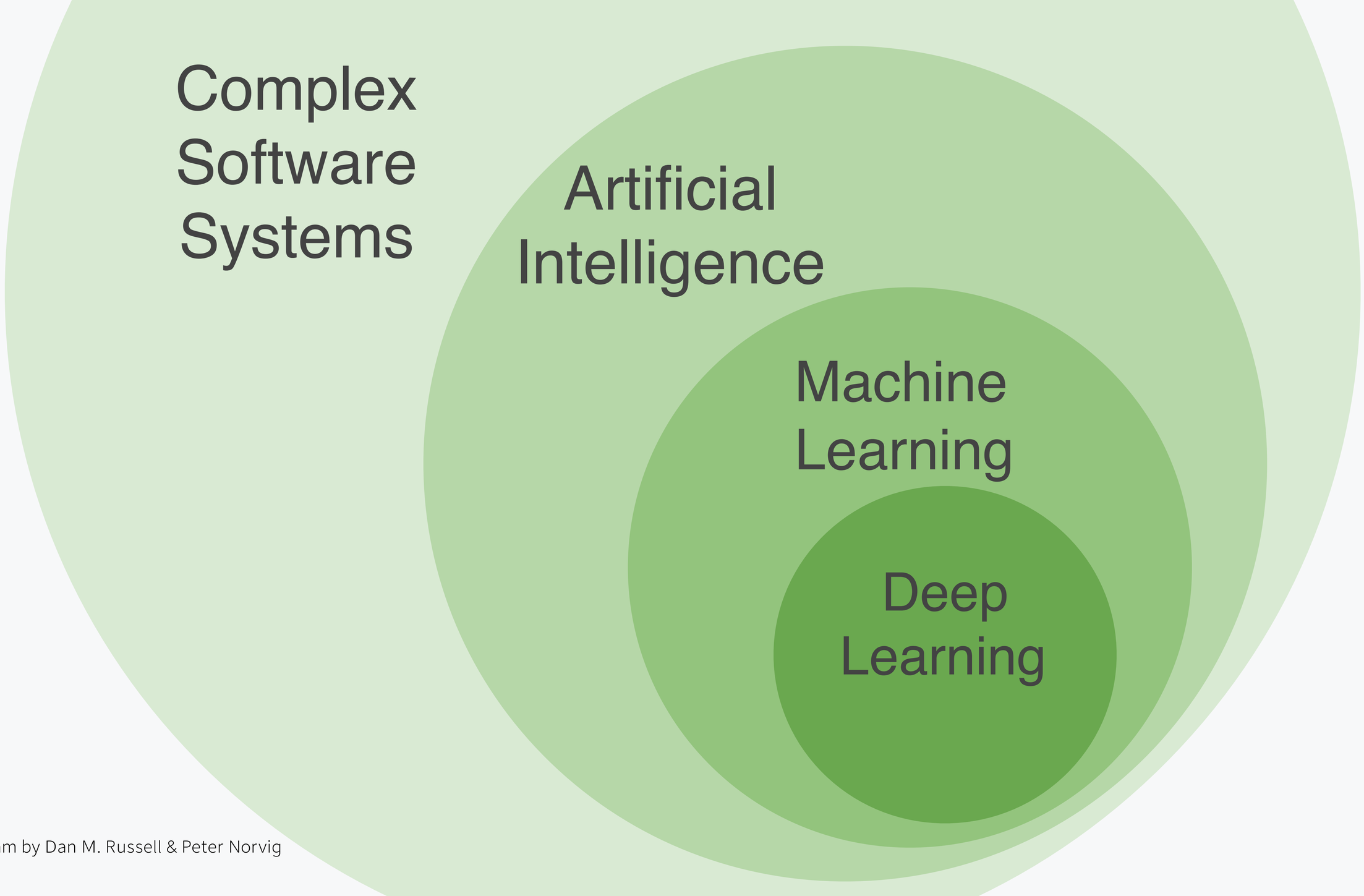
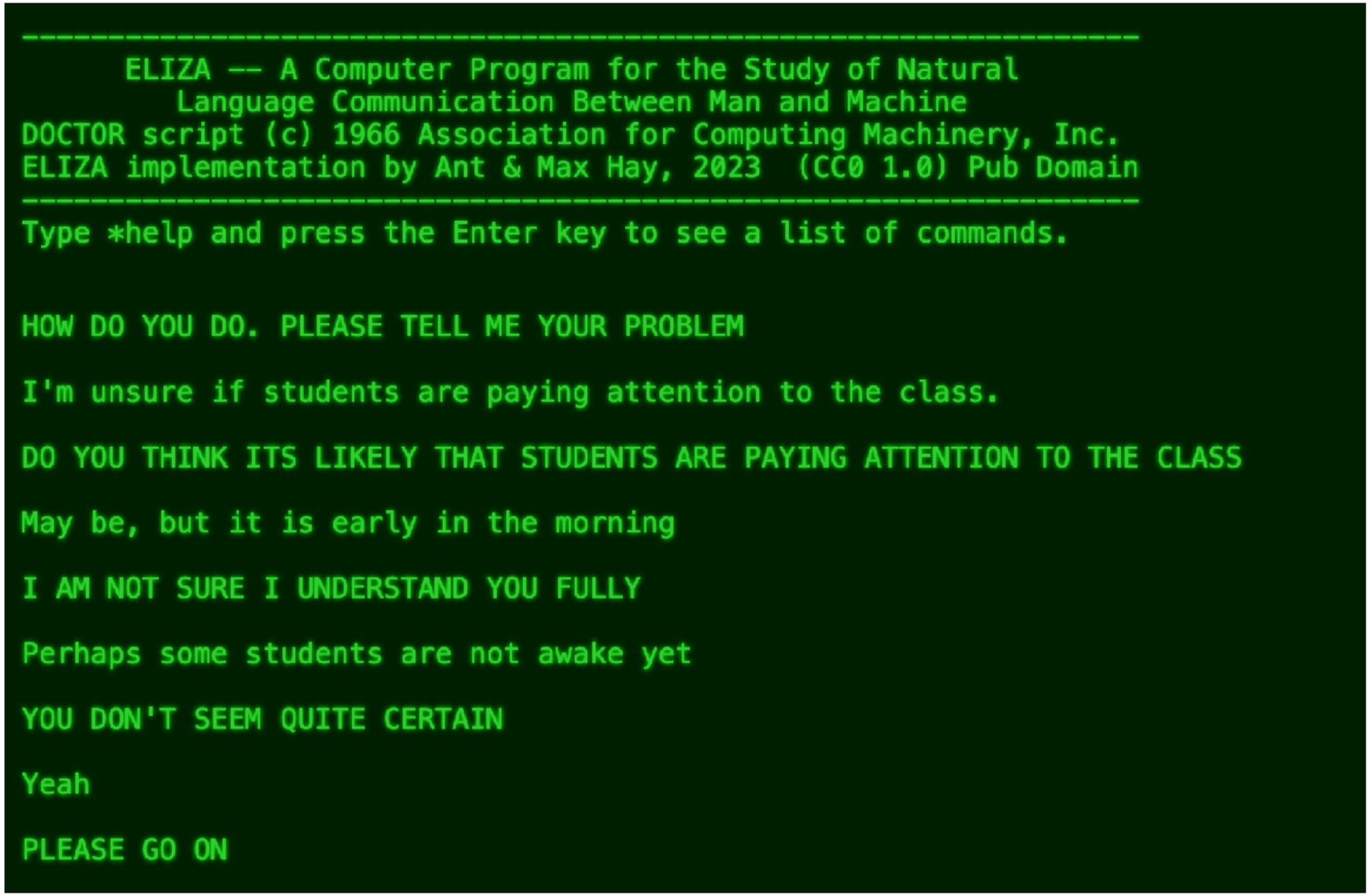


 Diagram by Dan M. Russell & Peter Norvig



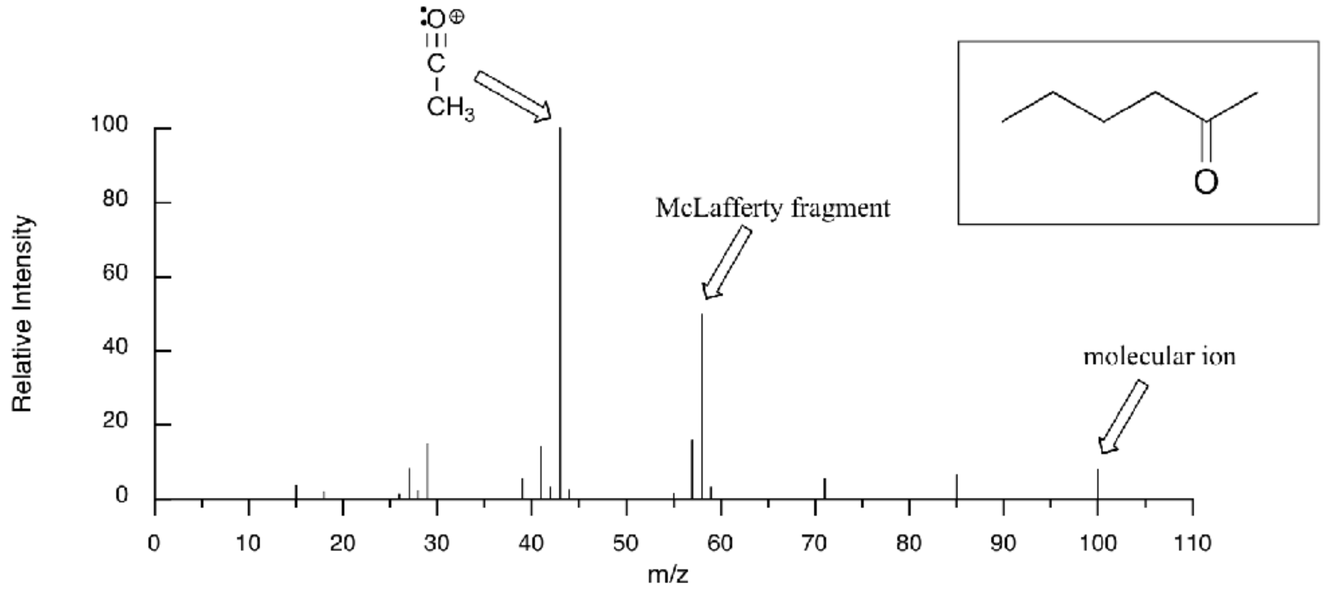
Eliza (1964–1967)

1. Identify keywords
2. If no keyword, use a preset response
3. Figure out the context
4. Choose transformations
5. Generate responses

 Genealogy of ELIZA • You can try ELIZA yourself at https://anthay.github.io/eliza.html

Dendral (1960–70):

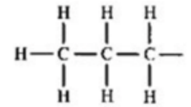
- Analyzes mass spectrometry
- Generate possible molecular structures
- Test them against chemistry knowledge base



IF

there is a high peak at 71amu
there is a high peak at 43amu
there is a high peak at 86amu
there is any peak at 58amu

THEN
there must be substructure:



Dendral history
Diagram by Tim Soderberg via LibreTexts

Deep Blue (1997): chess champ

Evaluate each move by searching a large database of play patterns that were trained by looking at grandmaster games

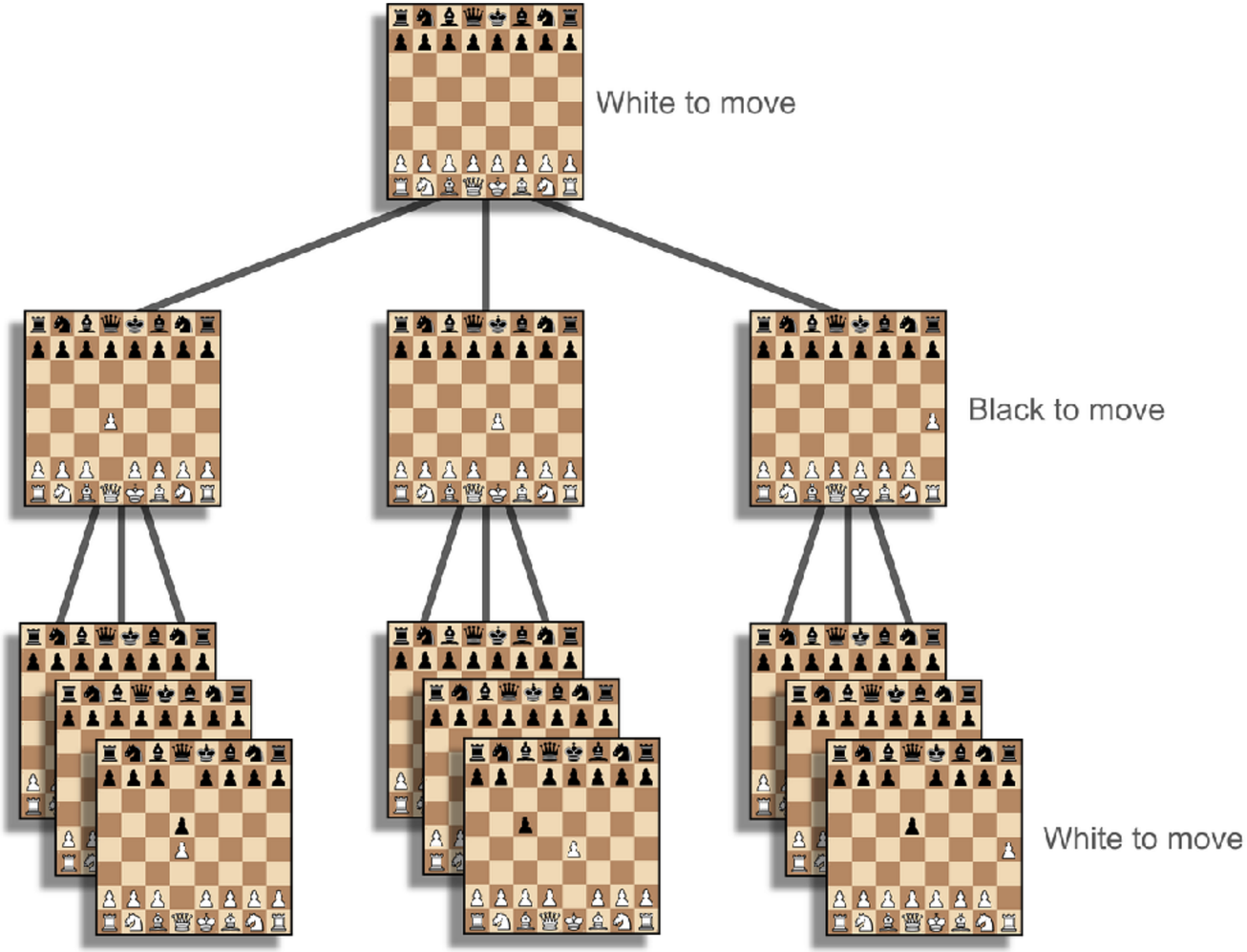


Photo:Kathy Willens/AP Images via Britannica
Diagram from Chess.com

Watson (2011): Jeopardy! champ

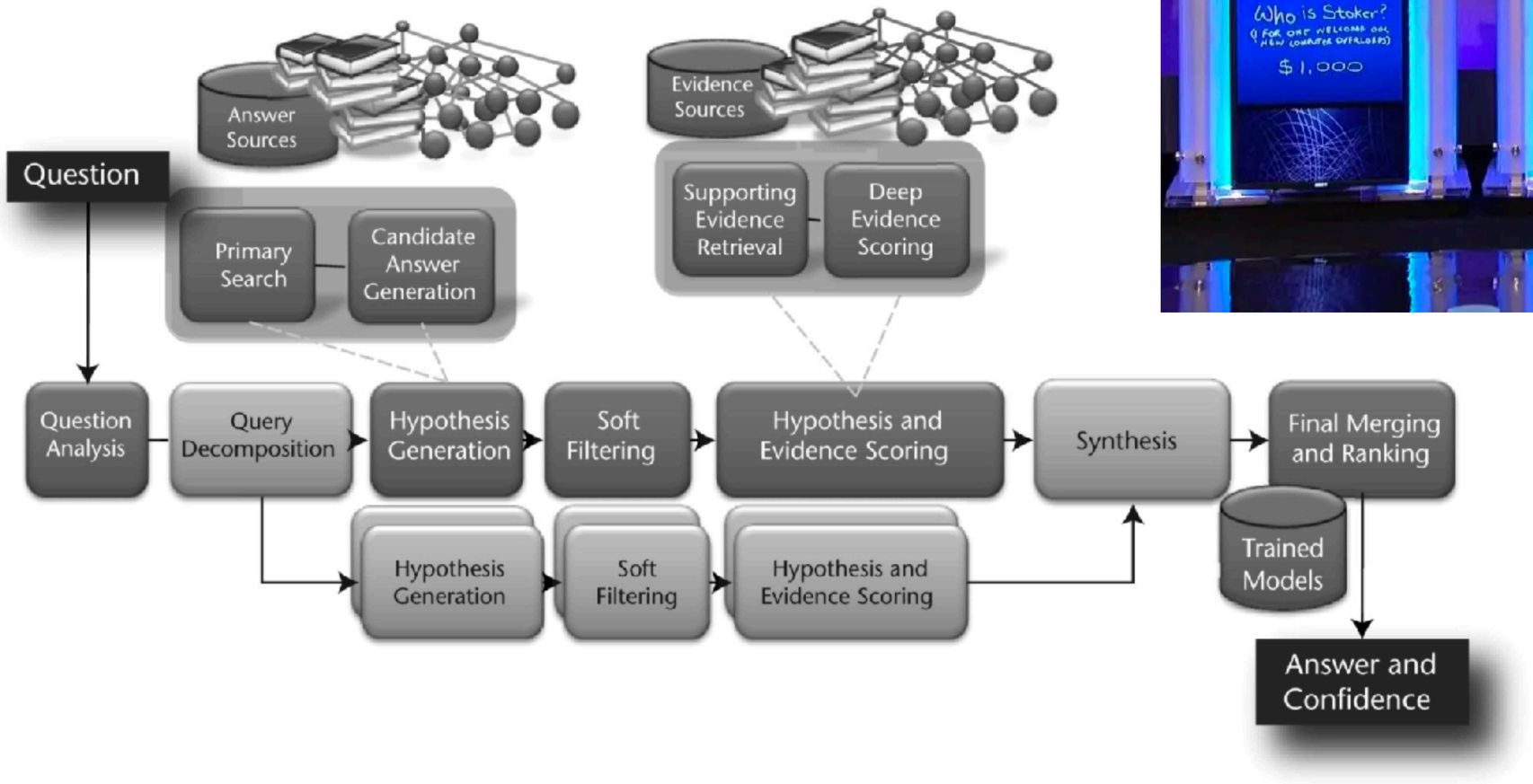
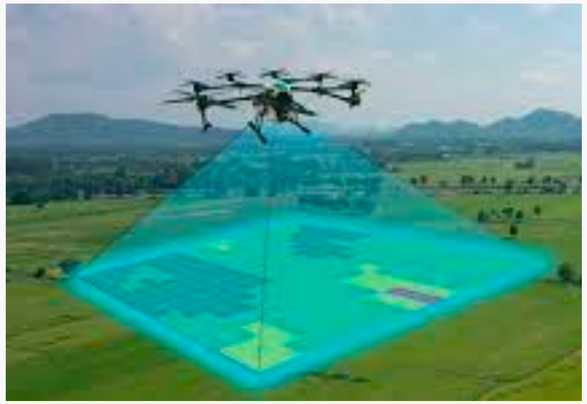


Diagram: Ferrucci et al. (2010) Building Watson
Photo: IBM

AI Systems Today

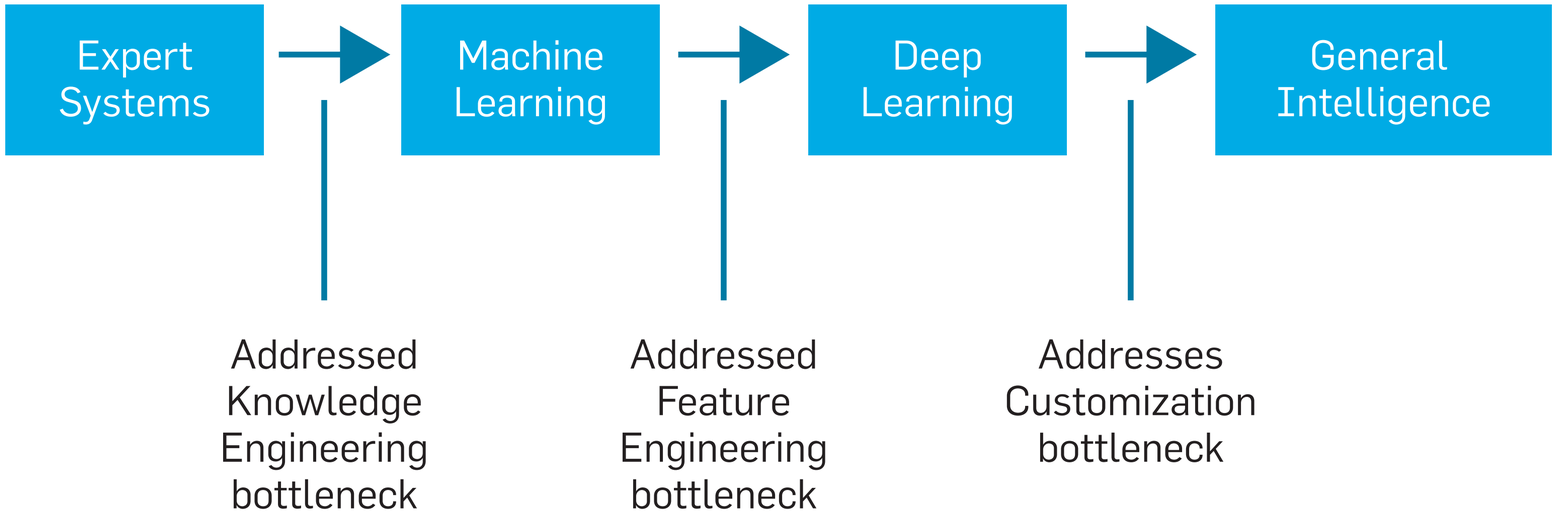


Title: Decline of Bee Populations and Its Impact on Global Agriculture: Causes, Consequences, and Solutions

Introduction

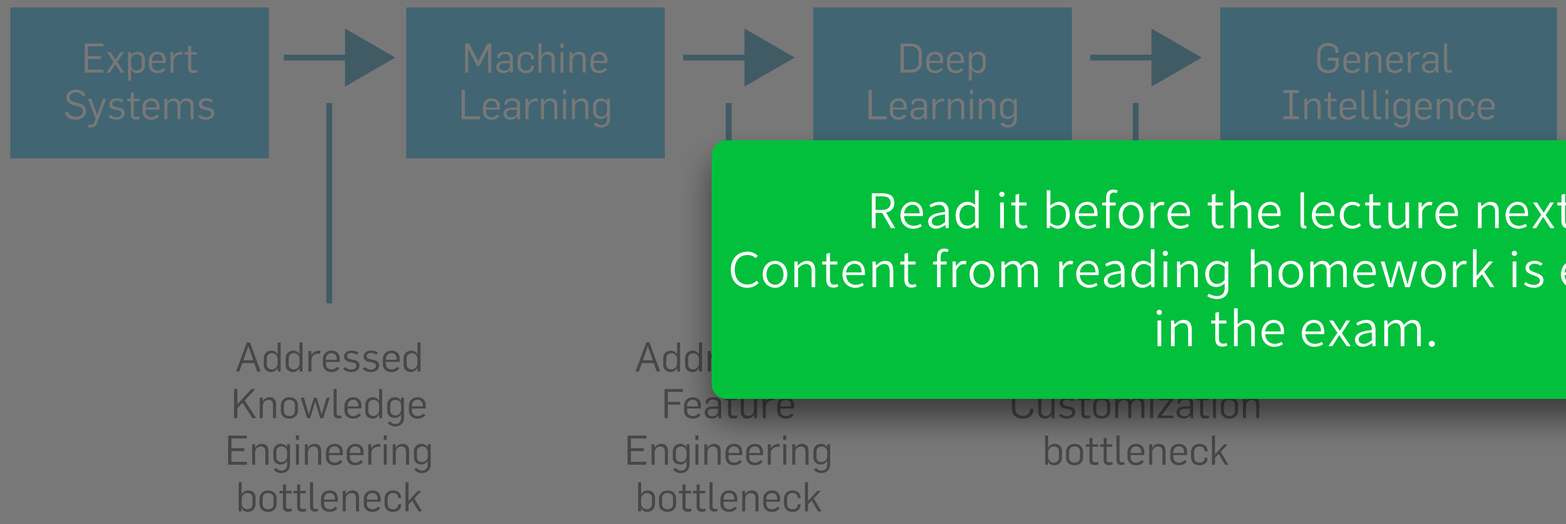
Bee populations worldwide have been experiencing a rapid decline in recent years, posing a significant threat to global agriculture. These vital pollinators contribute to the fertilization and reproduction of approximately 75% of the world's flowering plants, including around 35% of global food crops (Potts et al., 2016). The decline in bee populations has dire consequences for agriculture, food security, and the economy. In this article, we will explore the causes of this decline, the impact on global agriculture, and potential solutions to address this pressing issue.

reading homework



	Data	Exemplar	Scope	Curation
Expert Systems	Human	Rules	Follows	High
Machine Learning	+ Databases	Rules/networks	+ Discovers relationships	Medium
Deep Learning	+ Sensory	Deep neural networks	+ Senses relationships	Low
General Intelligence	+ Everything	Pre-trained deep neural networks	+ Understands the world	Minimal

 Dhar (2024) The Paradigm Shifts in Artificial Intelligence



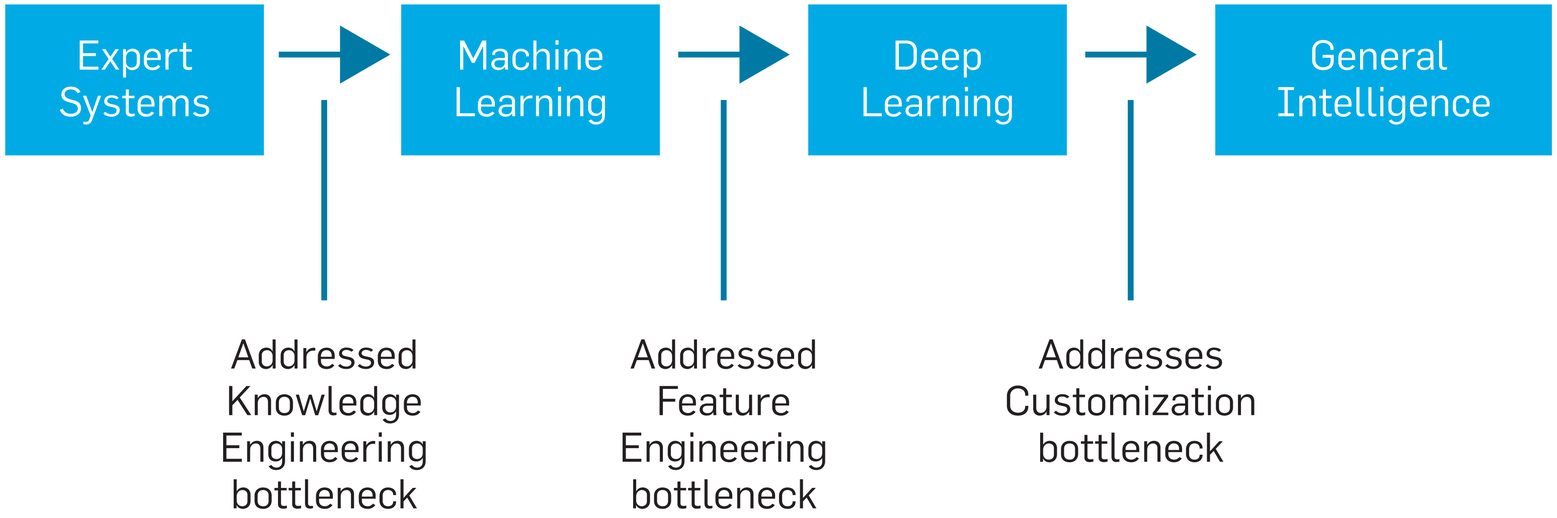
reading homework

Read it before the lecture next week.
Content from reading homework is eligible to be in the exam.

	Data	Exemplar	Scope	Curation
Expert Systems	Human	Rules	Follows	High
Machine Learning	+ Databases	Rules/networks	+ Discovers relationships	Medium
Deep Learning	+ Sensory	Deep neural networks	+ Senses relationships	Low
General Intelligence	+ Everything	Pre-trained deep neural networks	+ Understands the world	Minimal

 Dhar (2024) The Paradigm Shifts in Artificial Intelligence

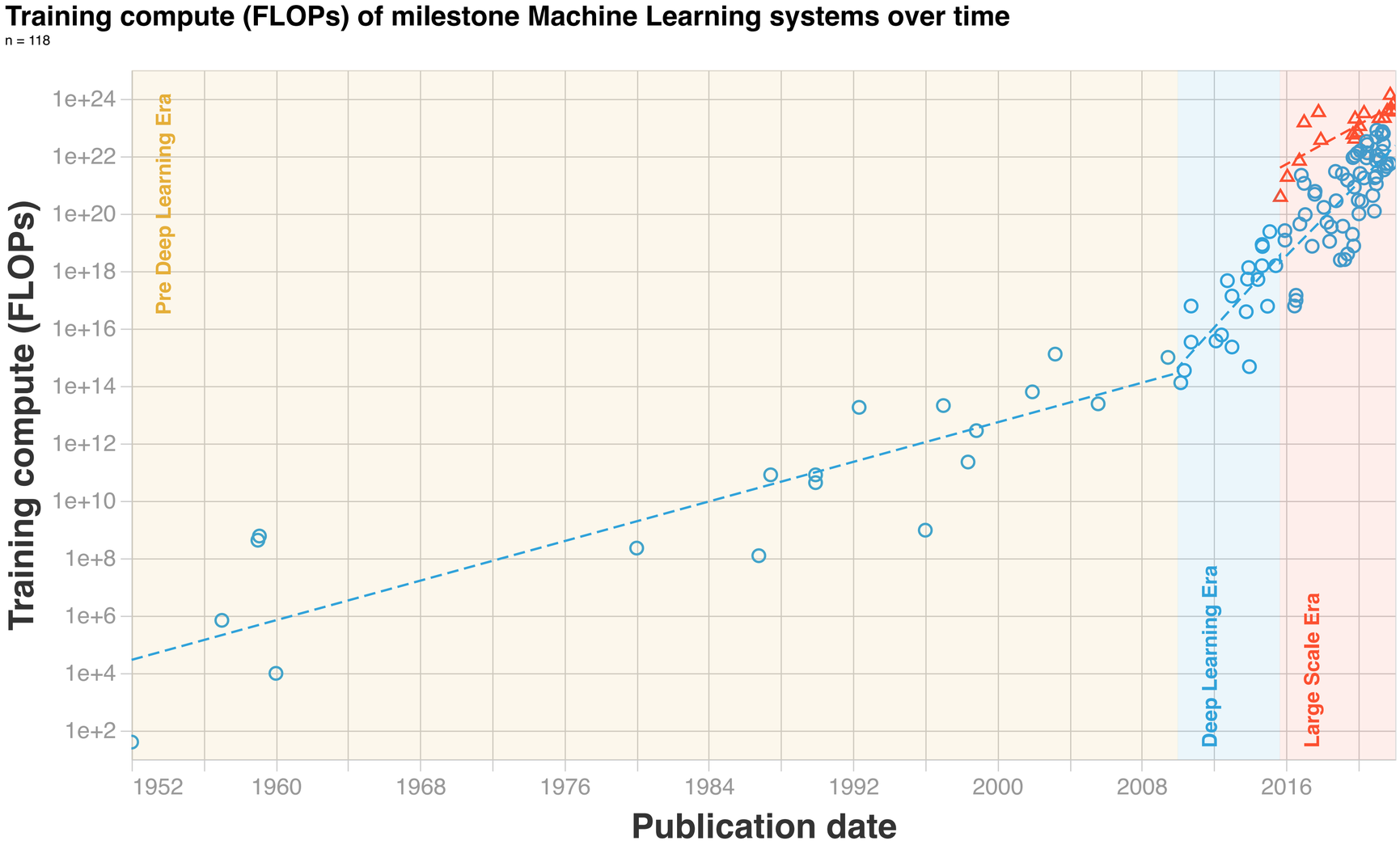
reading homework



	Data	Exemplar	Scope	Curation
Expert Systems	Human	Rules	Follows	High
Machine Learning	+ Databases	Rules/networks	+ Discovers relationships	Medium
Deep Learning	+ Sensory	Deep neural networks	+ Senses relationships	Low
General Intelligence	+ Everything	Pre-trained deep neural networks	+ Understands the world	Minimal

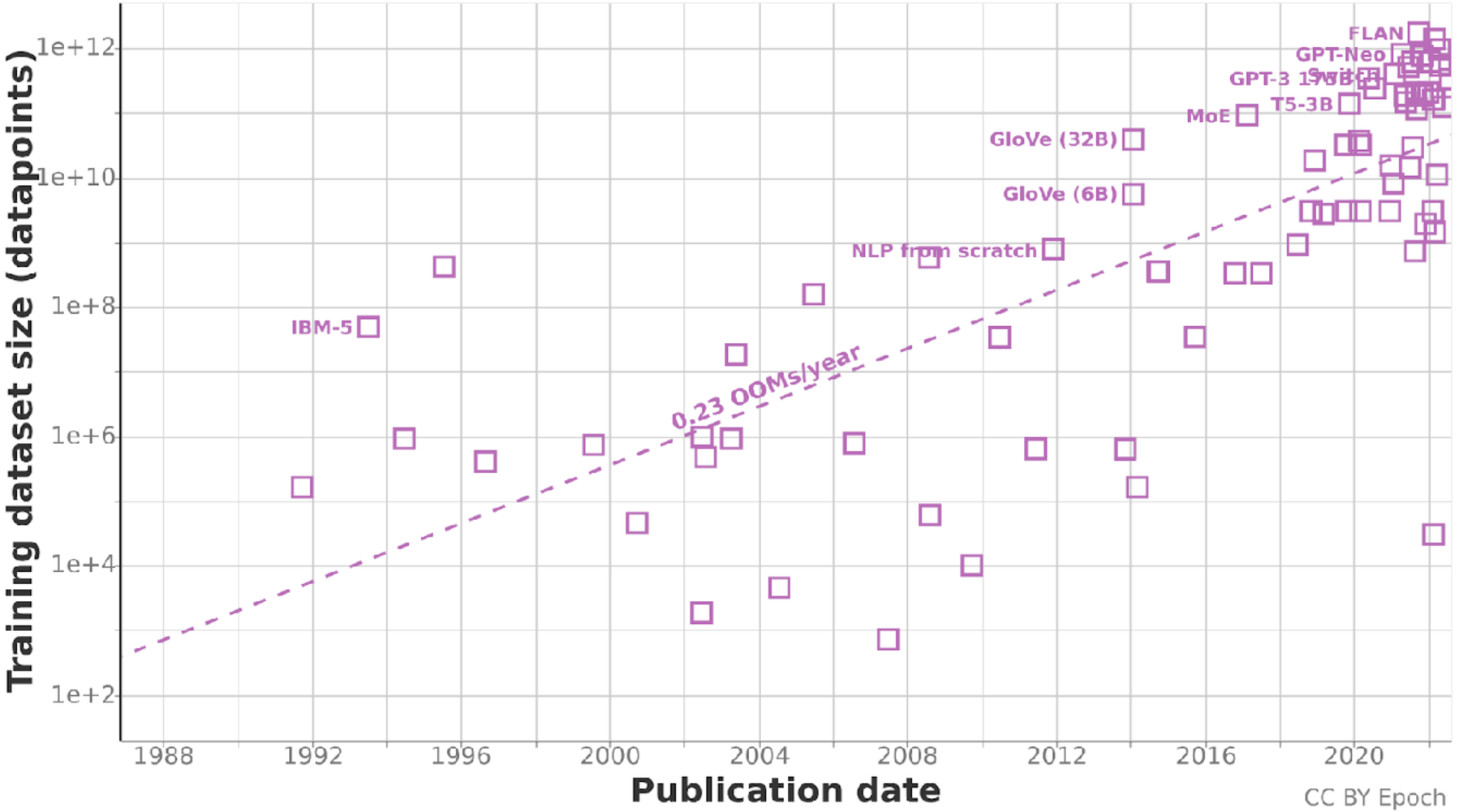
 Dhar (2024) The Paradigm Shifts in Artificial Intelligence

Explosion in computation



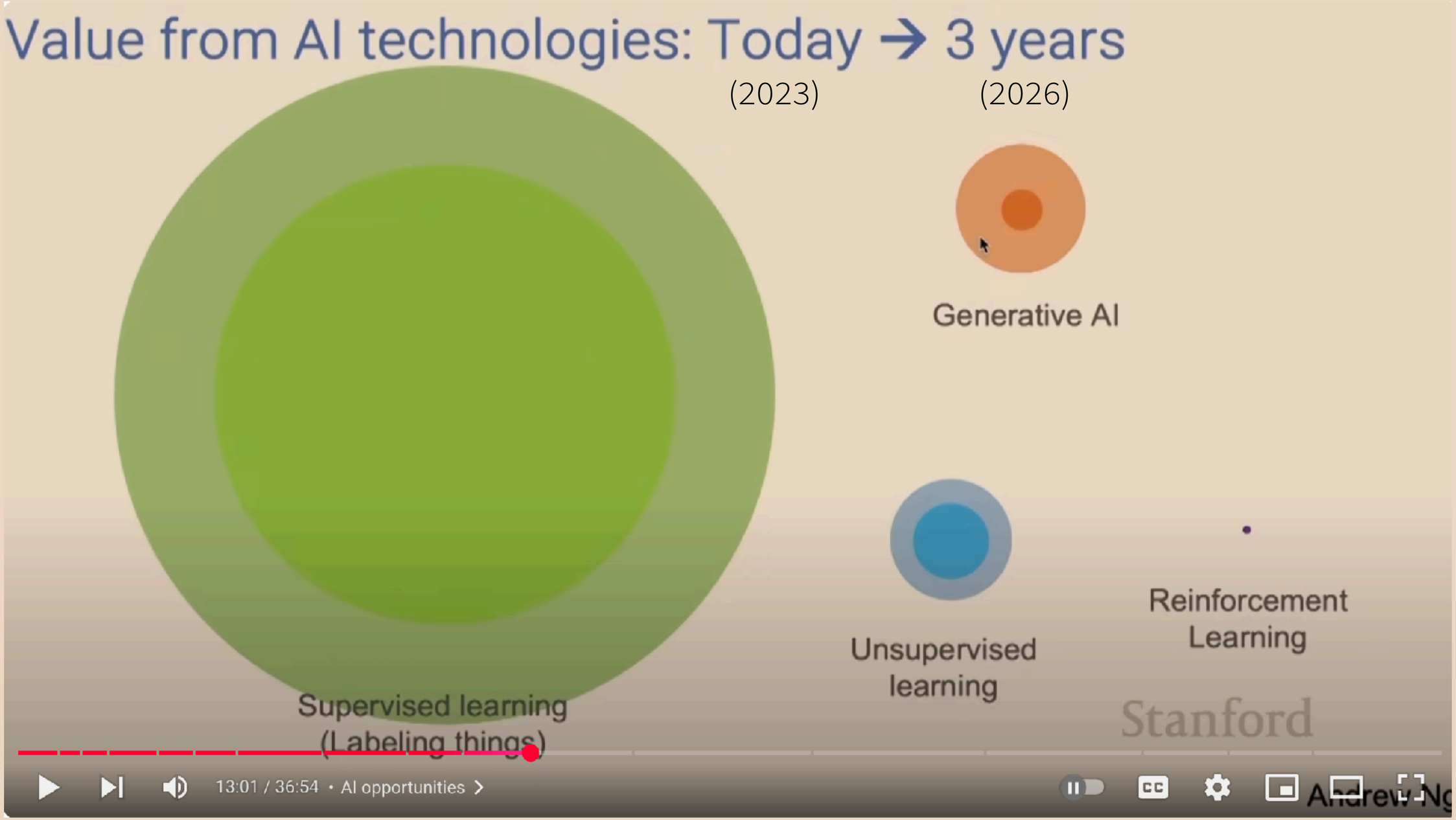
 Sevilla et al. (2022) [Compute Trends Across Three Eras of Machine Learning]

Explosion in training data size

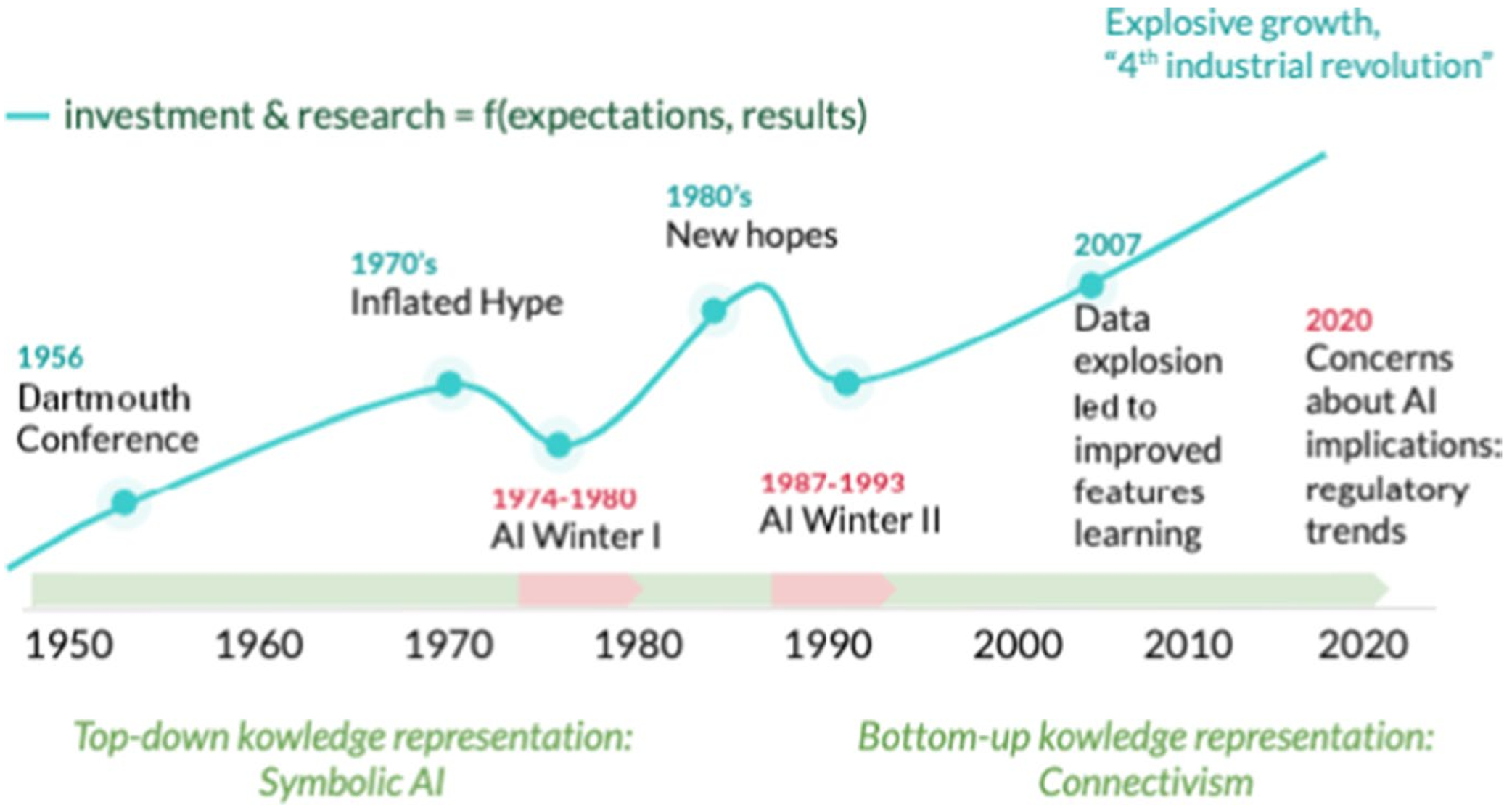


 Pablo Villalobos and Anson Ho (2022), Trends in Training Dataset Sizes (Published online at epoch.ai)

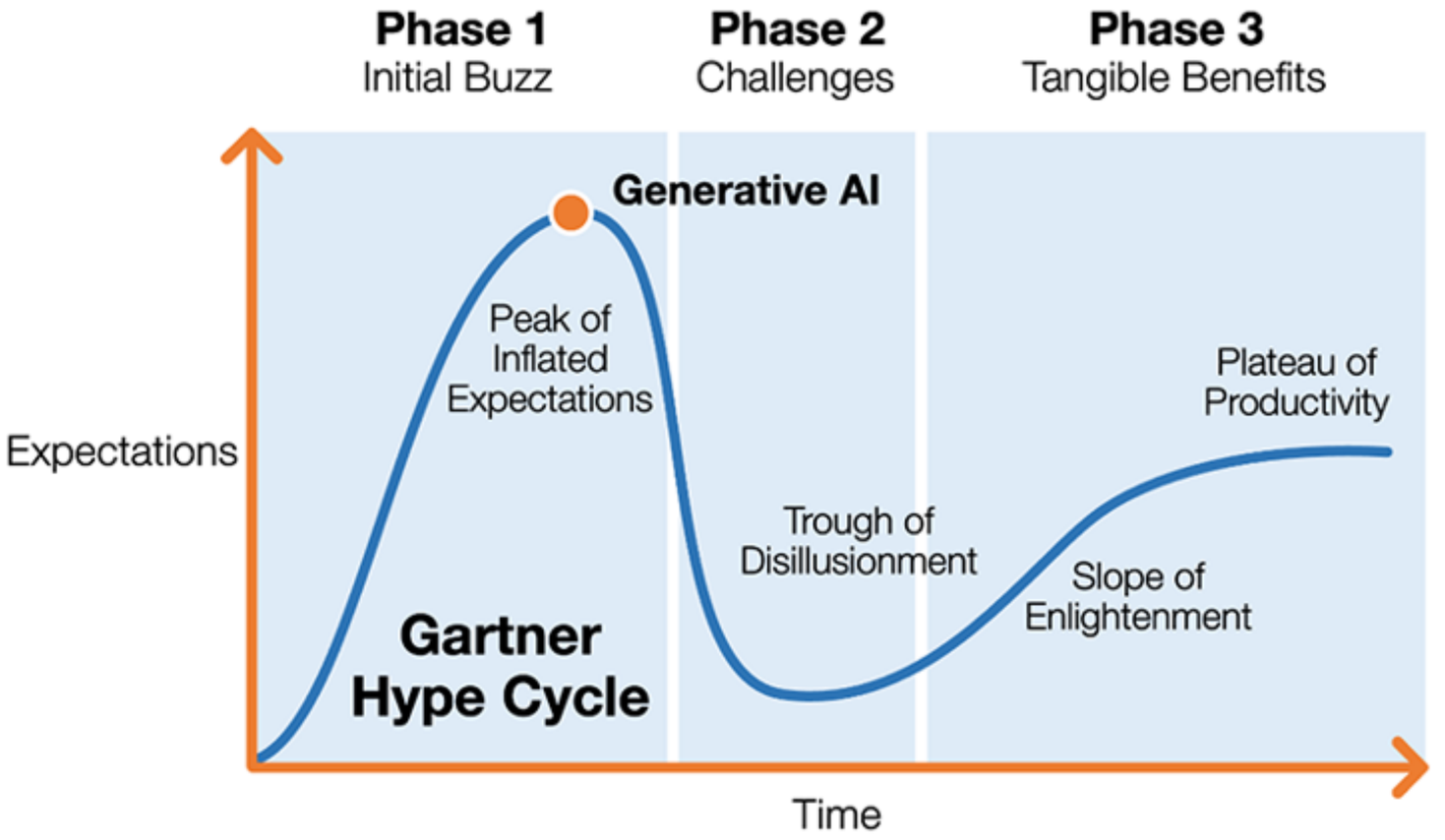
 Andrew Ng (2023) Opportunities in AI on YouTube



 Andrew Ng (2023) Opportunities in AI on YouTube



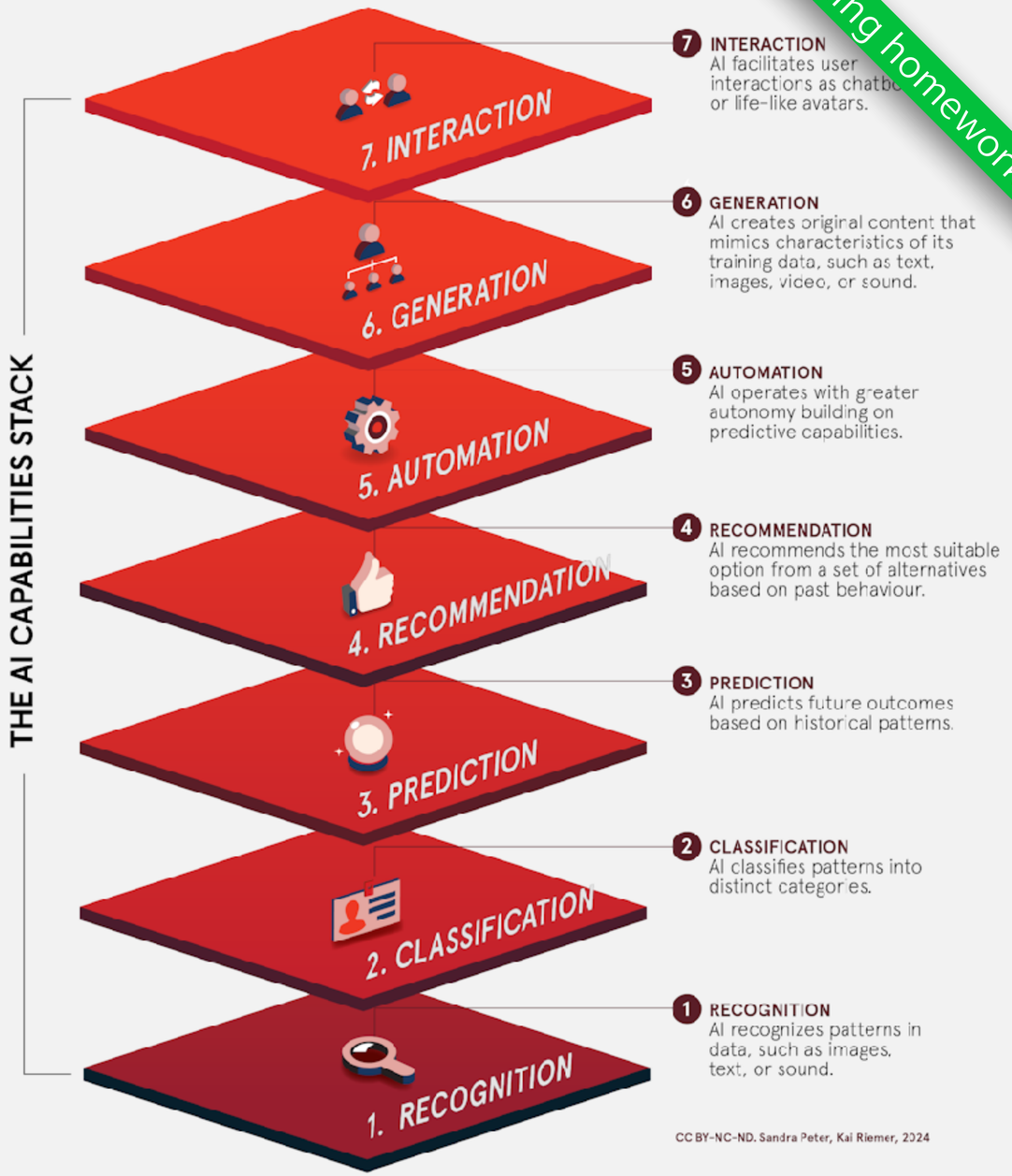
 Francesconi (2022) [The winter, the summer and the summer dream of artificial intelligence in law]



31 Diagram: Brent Dykes from Gartner via Forbes

reading homework

Capabilities of AI systems



Peter & Riemer (2024) Wondering what AI actually is? Here are the 7 things it can do for you (an online article on The Conversation website)

Exercise: How search works

⏳ 10 minutes

1. Take a piece of paper and draw how Google Search works — according to your understanding

2. Turn to your neighbor and explain to each other what you drew

Different models of how a system works

Designer's conceptual model — conception of the look, feel, and operation of a product. The system image is what can be derived from the structure and operation.

User's mental model developed by interaction with the product and by extrapolation from previous, similar systems

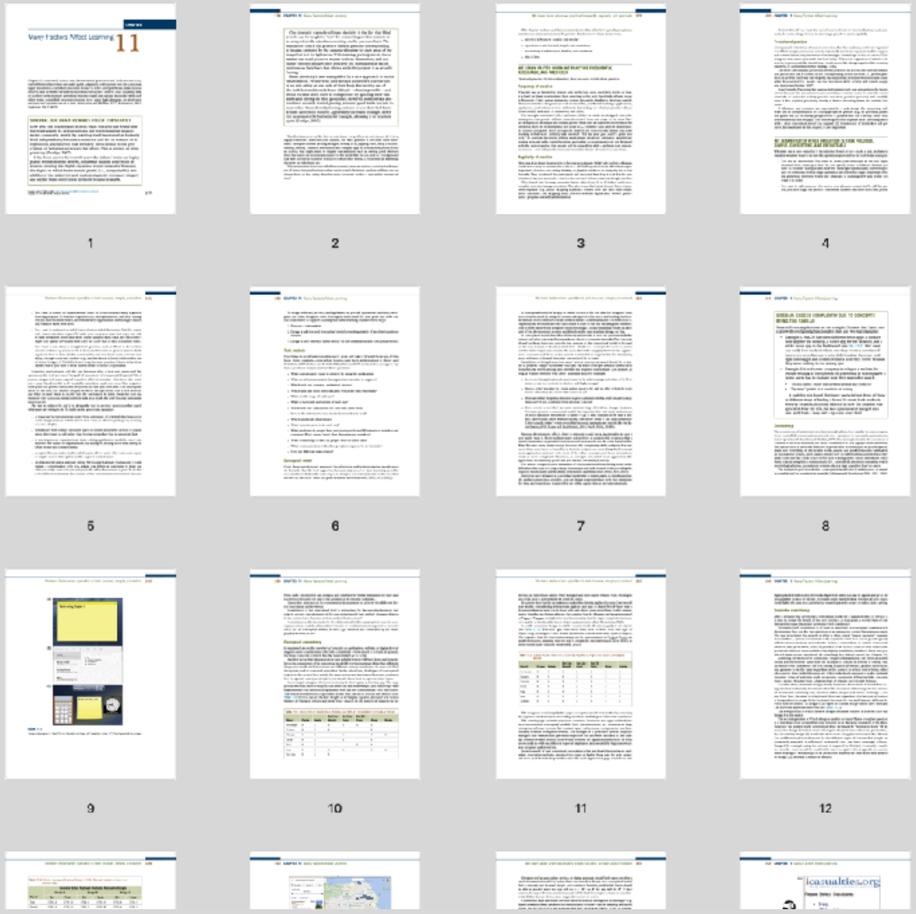
Designers expect the user's model to be identical to their own, but because they cannot communicate directly with the user, the burden of communication is through the design

Mental models are usually incomplete.
They may be inaccurate.
They are updated as little as possible.

 Diagram: Norman (2013) The Design of Everyday Things: Revised and Expanded Edition (p. 32)

Johnson (2020) Ch. 11: Many factors affect learning

reading homework



- Some psychology of learning that are used in designing (non-AI) interactive systems
- Description of users' mental model and designer's conceptual model
- Questions to consider when analyzing users' tasks

Homework

- Chapter 11 from Johnson (2020) Designing with the Mind in Mind (3 ed.)
  - Focus on the following sections on pp. 179 – 199
  - We learn faster when operation is task focused, simple, consistent, and predictable
  - When risk is low, we explore more and learn more
- Dhar (2024) The Paradigm Shifts in AI
- Peter & Riemer (2024) Wondering what AI actually is? Here are the 7 things it can do for you

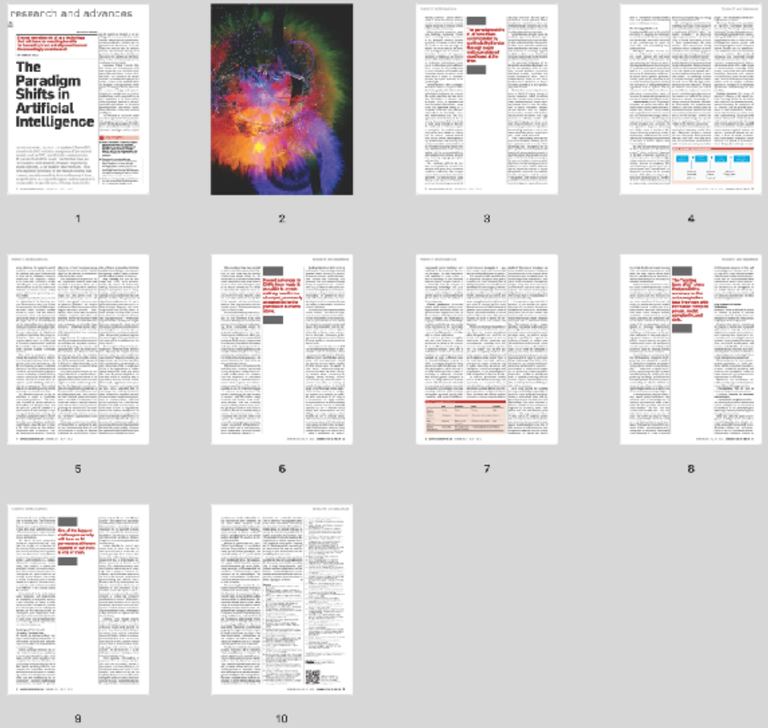
- Read the course syllabus and note down any questions you may have

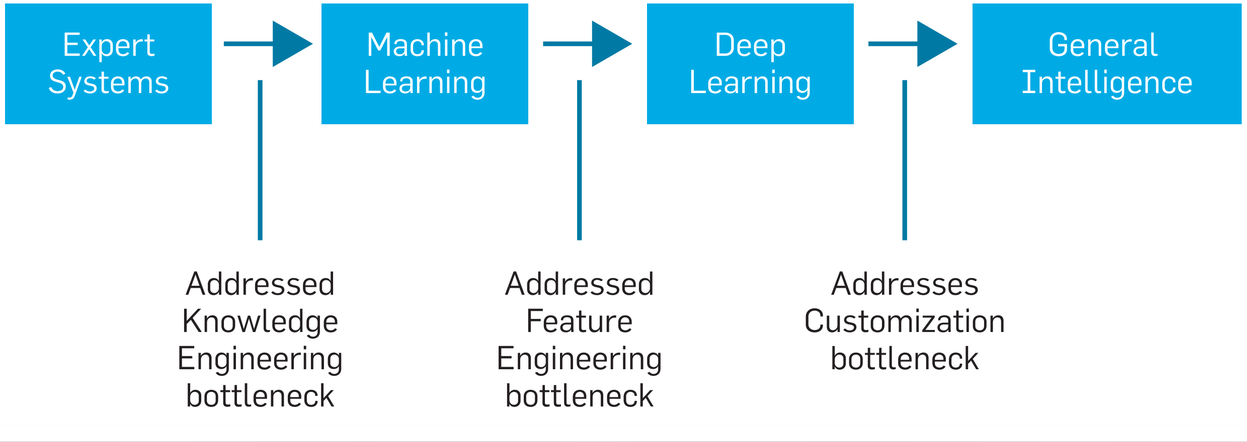
📗 **Read before the lecture next week**

❗ Next lecture will be on Thursday, September 25 (no lecture on Wednesday)

Dhar (2024) The Paradigm Shifts in AI

reading homework

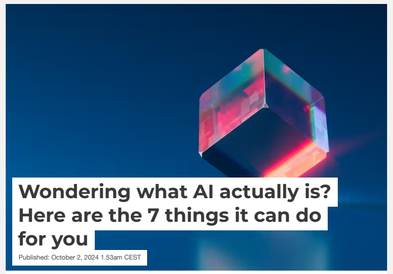




	Data	Exemplar	Scope	Curation
Expert Systems	Human	Rules	Follows	High
Machine Learning	+ Databases	Rules/networks	+ Discovers relationships	Medium
Deep Learning	+ Sensory	Deep neural networks	+ Senses relationships	Low
General Intelligence	+ Everything	Pre-trained deep neural networks	+ Understands the world	Minimal

Peter & Riemer (2024) AI capabilities

reading homework



Dacachi/Shutterstock

You know we've reached peak interest in artificial intelligence (AI) when Oprah Winfrey hosts a television special about it. AI is truly everywhere. And we will all have a relationship with it – whether using it, building it, governing it or even befriending it.

But what exactly is AI? While most people won't need to know exactly how it works under the hood, we will all need to understand what it can do. In our conversations with global leaders across business, government and the arts, one thing stood out – you can't fake it anymore. AI fluency that is.

AI isn't just about chatbots. To help understand what it is about, we've developed a framework which explains the broad broad range of capabilities it offers. We call this the "capabilities stack".

We see AI systems as having seven basic kinds of capability, each building on the ones below it in the stack. From least complex to most, these are: recognition, classification, prediction, recommendation, automation, generation and interaction.

Recognition

At its core, the kind of AI we are seeing in consumer products today identifies patterns. Unlike traditional coding, where developers explicitly program how a system works, AI "learns" these patterns from vast datasets, enabling it to perform tasks. This "learning" is essentially just advanced mathematics that turns patterns into complex probabilistic models – encoded in so-called artificial neural networks.

Once learned, patterns can be recognised – such as your face, when you open your phone, or when you clear customs at the airport.

Classification

Once an AI system can recognise patterns, we can train it to detect subtle variations and categorise them. This is how your photo app neatly organises albums by family members, or how apps identify and label different kinds of skin lesions. AI classification is also at work behind the scenes when phone companies and banks identify spam and fraud calls.

In New Zealand, non-profit organisation Te Hiku developed an AI language model to classify thousands of hours of recordings to help revitalise Te Reo Māori, the local indigenous language.

Prediction

When AI is trained on past data, it can be used to predict future outcomes. For example, airlines use AI to predict the estimated arrival times of incoming flights and to assign gates on time so you don't end up waiting on the tarmac.

Similarly, Google Flights uses AI to predict flight delays even before airlines announce them.

In Hong Kong, an AI prediction model saves taxpayer money by predicting when a project needs early intervention to prevent it overrunning its budget and completion date. And when you buy stuff on Amazon, the ecommerce giant uses AI to predict demand and optimise delivery routes, so you get your packages within hours, not just days.

Recommendation

Once we predict, we can make recommendations for what to do next.

If you went to Taylor Swift's Eras tour concert at Sydney's Accor stadium, you were kept safe thanks to AI recommendations. A system funded by the New South Wales government used data from multiple sources to analyse the movement and mood of the 80,000 strong crowd, providing real-time recommendations to ensure everyone's safety.

AI-based recommendations are everywhere. Social media, streaming platforms, delivery services and shopping apps all use past behaviour patterns to present you with their "for you" pages. Even pig farms use pig facial recognition and tracking to alert farmers to any issues and recommend particular interventions.

Automation

It's a small step from prediction and recommendation to full automation.

In Germany, large wind turbines use AI to keep the lesser spotted eagle safe. An AI algorithm detects approaching birds and automatically slows down the turbines allowing them to pass unharmed.

Closer to home, Melbourne Water uses AI to autonomously regulate its pump control system to reduce energy costs by around 20% per year. In Western Sydney, local buses on key routes are AI-enabled: if a bus is running late, the system predicts its arrival at the next intersection and automatically green-lights its journey.

Generation

Once we can encode complex patterns into neural networks, we can also use these patterns to generate new, similar ones. This works with all kinds of data – images, text, audio and video.

Image generation is now built into many new phones. Don't like the look on someone's face? Change into a smile. Want a boat on that lake? Just add it in. And it doesn't stop there.

Tools such as Runway let you manipulate videos or create new ones with just a text prompt. ElevenLabs allows you to generate synthetic voices or digitise existing ones from short recordings. These can be used to narrate audiobooks, but also carry risks such as deepfake impersonation.

And we haven't even mentioned large language models such as ChatGPT, which are transforming how we work with text and how we develop computer code. Research by McKinsey found that these models can cut the time required for complex coding tasks by up to 50%.

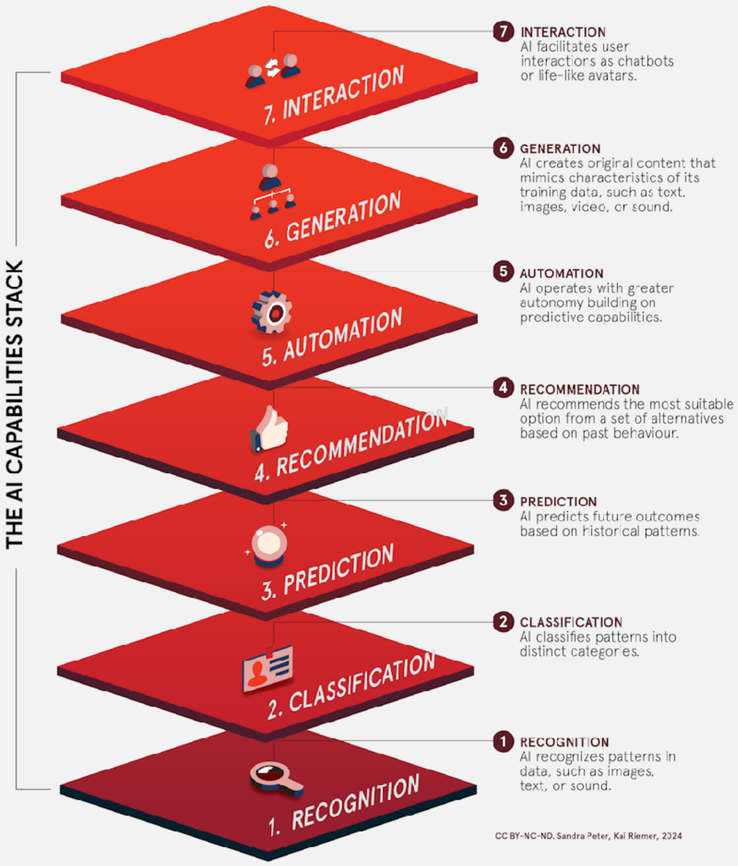
Interaction

Finally, generative AI also makes it possible to mimic human-like interactions.

Soon, virtual assistants, companions and digital humans will be everywhere. They will attend your Zoom meeting to take notes and schedule follow-up meetings.

Interactive AI assistants, such as IBM's AskHR bot, will answer your HR questions. And when you get home, your AI friend app will entertain you, while digital humans on social media are ready to sell you anything, any time. And with voice mode activated, even ChatGPT gets in on the inter-action.

Amid the excitement around generative AI, it is important to remember that AI is more than chatbots. It impacts many things beyond the flashy conversational tools – often in ways that quietly improve everyday processes.



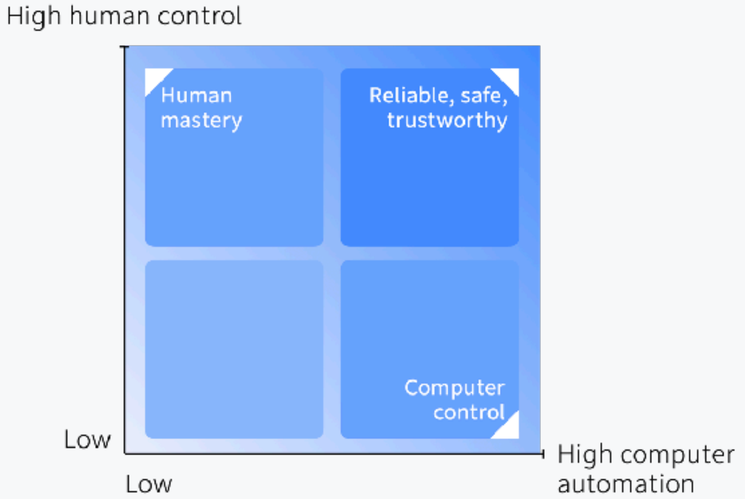
Debrief question

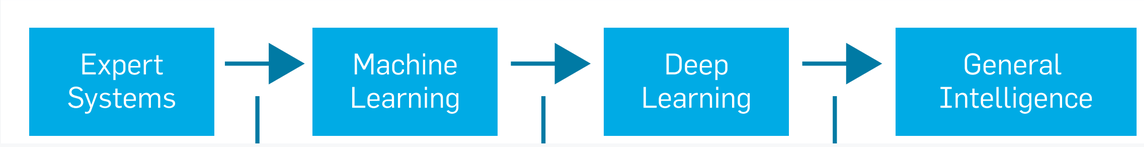
When you design an AI system, why is it useful to analyze the level of control × automation?

https://chatw.ch/hcai25

Under **Lecture 2** section,
click **Debrief question**

When you are done, you may leave

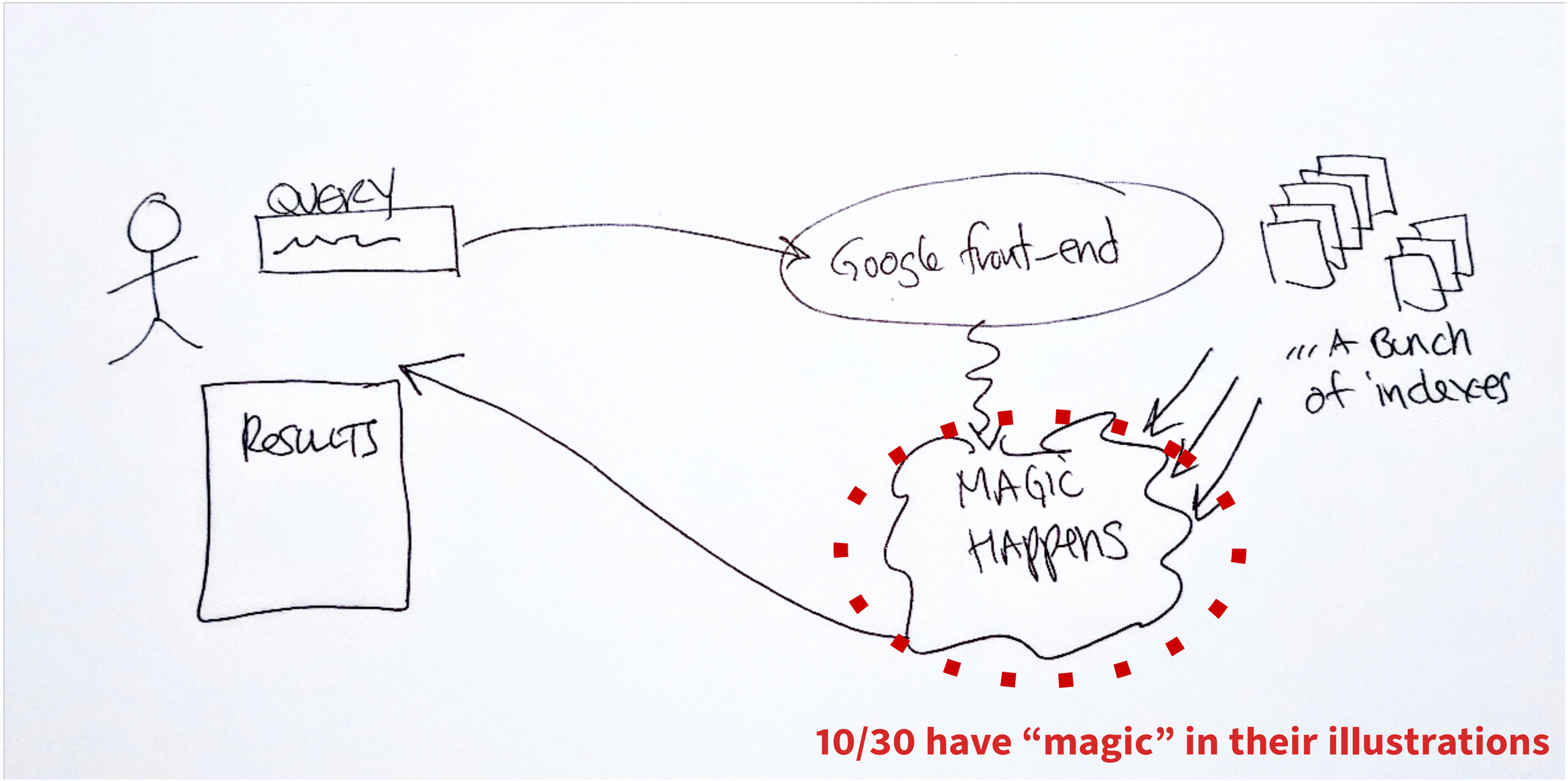




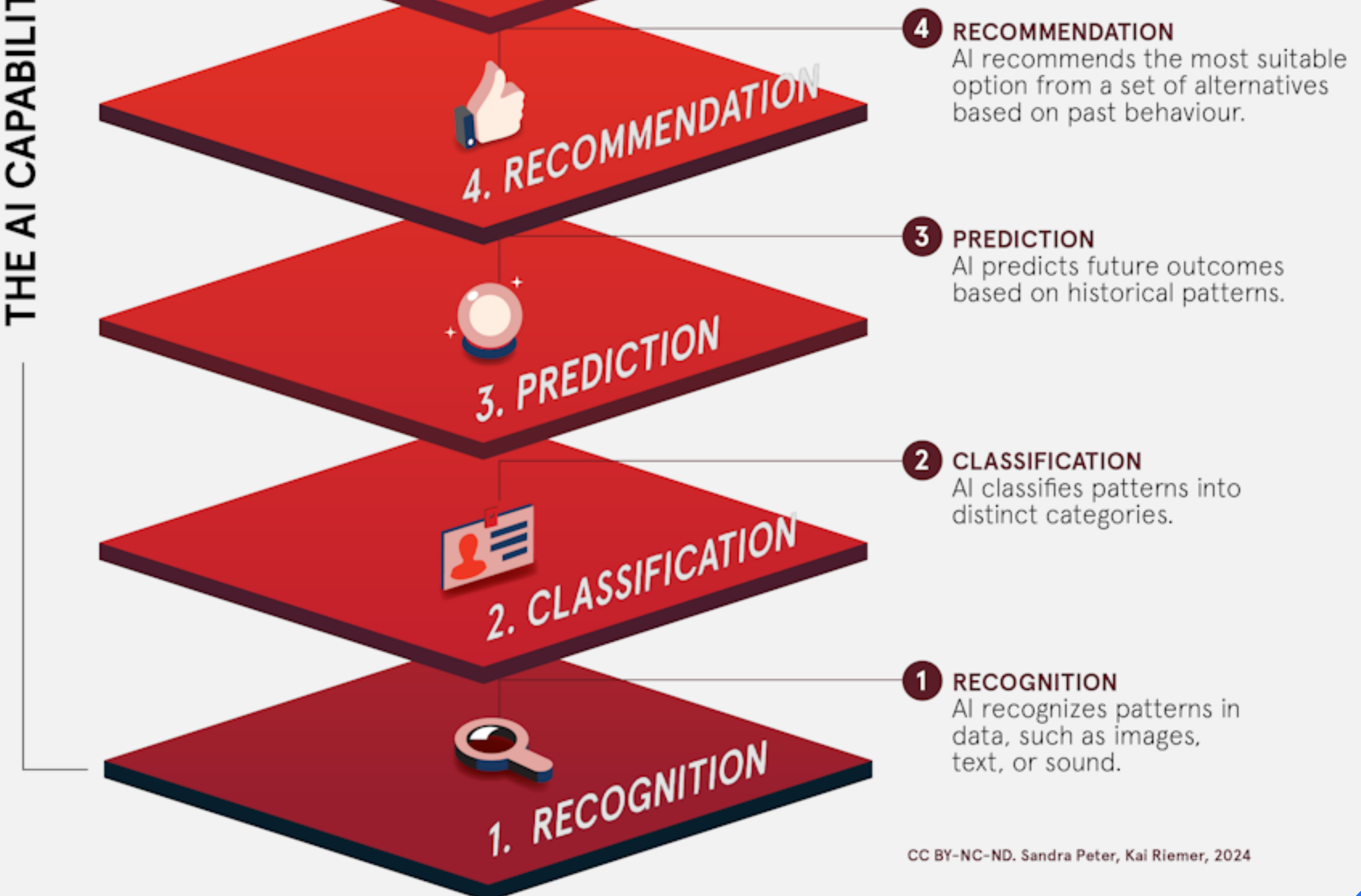
❗ Next lecture will be on Thursday, September 25 (no lecture on Wednesday)

Supplemental slides

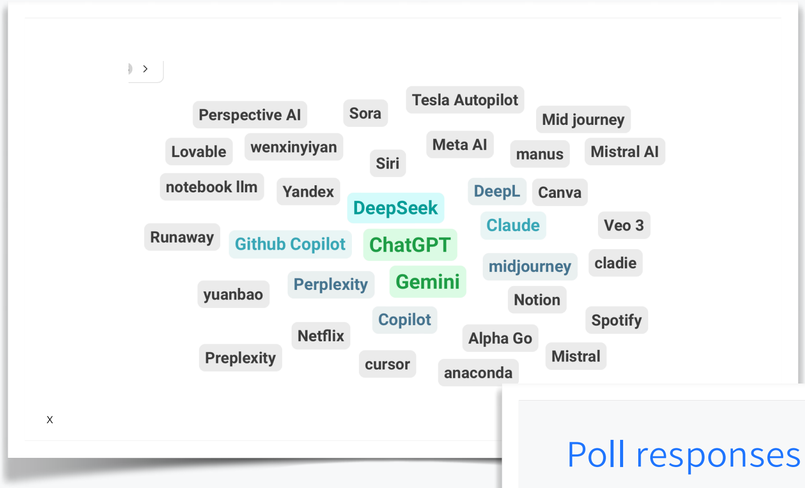
One student's model of search

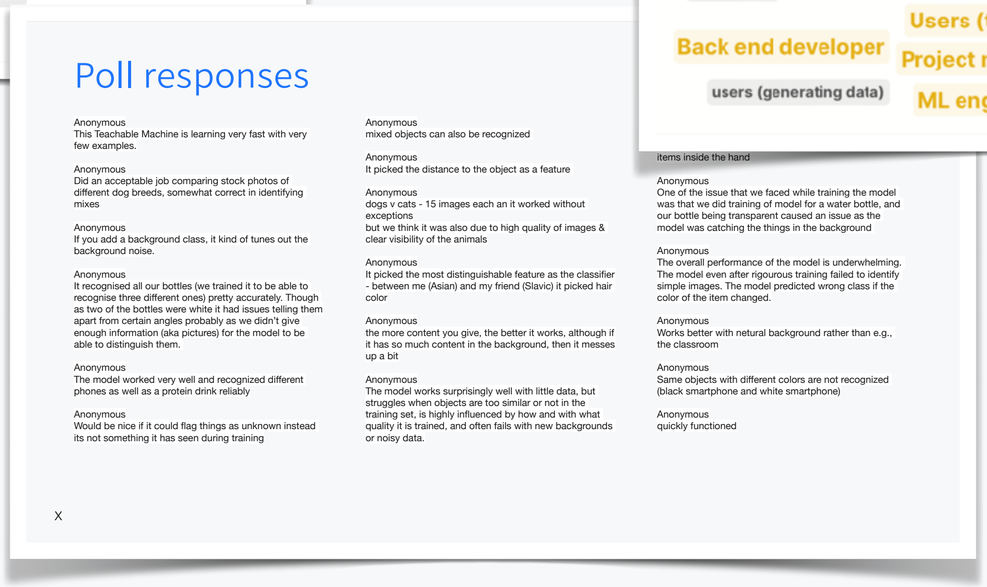


10/30 have "magic" in their illustrations



Your work from yesterday were added to OLAT





Poll responses

Anonymous
This Teachable Machine is learning very fast with very few examples.

Anonymous
Did an acceptable job comparing stock photos of different dog breeds, somewhat correct in identifying mixes

Anonymous
If you add a background class, it kind of tunes out the background noise.

Anonymous
It recognised all our bottles (we trained it to be able to recognise three different ones) pretty accurately. Though as two of the bottles were white it had issues telling them apart from certain angles probably as we didn't give enough information (aka pictures) for the model to be able to distinguish them.

Anonymous
The model worked very well and recognized different phones as well as a protein drink reliably

Anonymous
Would be nice if it could flag things as unknown instead its not something it has seen during training

Anonymous
mixed objects can also be recognized

Anonymous
It picked the distance to the object as a feature

Anonymous
dogs v cats - 15 images each an it worked without exceptions
but we think it was also due to high quality of images & clear visibility of the animals

Anonymous
It picked the most distinguishable feature as the classifier - between me (Asian) and my friend (Slavic) it picked hair color

Anonymous
the more content you give, the better it works, although if it has so much content in the background, then it messes up a bit

Anonymous
The model works surprisingly well with little data, but struggles when objects are too similar or not in the training set, is highly influenced by how and with what quality it is trained, and often fails with new backgrounds or noisy data.

items inside the hand

Anonymous
One of the issue that we faced while training the model was that we did training of model for a water bottle, and our bottle being transparent caused an issue as the model was catching the things in the background

Anonymous
The overall performance of the model is underwhelming. The model even after rigourous training failed to identify simple images. The model predicted wrong class if the color of the item changed.

Anonymous
Works better with netural background rather than e.g., the classroom

Anonymous
Same objects with different colors are not recognized (black smartphone and white smartphone)

Anonymous
quickly functioned

https://chatw.ch/hcai25

(Slides for today will be available around the middle of this lecture.)